

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:28:32 : Search time 536 Seconds
(Without alignments)
8530.060 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043
Sequence: 1 gaattccattgtgttgta.....tgggacccaataaaaaaa 2043

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapept 60.0

Searched: 1504479 seqs, 1118970152 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2975614

Minimum DB seq length: 15

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published Applications_NA:*

1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PTCT_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpna/PTCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1547	75.7	2032	9 US-09-816-825-1	Sequence 1, Appl1
2	1547	75.7	2032	13 US-10-007-262-2	Sequence 2, Appl1
3	403	19.7	505	10 US-09-998-598-2595	Sequence 2595, Ap
4	403	19.7	517	10 US-09-919-580-194	Sequence 194, Ap
5	114	5.6	389	10 US-09-919-580-264	Sequence 264, App
6	59	2.9	2544	9 US-09-927-602-1	Sequence 1, Appl1
7	59	2.9	48436	9 US-09-927-602-8	Sequence 38, Appl1
8	37	1.8	37	9 US-09-816-825-7	Sequence 7, Appl1
9	37	1.8	13	13 US-10-007-262-7	Sequence 7, Appl1
10	22	1.1	167	10 US-09-867-701-9458	Sequence 9458, Ap
11	22	1.1	171	10 US-09-867-701-9067	Sequence 9067, Ap
12	22	1.1	187	10 US-09-867-701-9185	Sequence 9185, Ap
13	22	1.1	191	10 US-09-867-701-9284	Sequence 9284, Ap
14	20	1.0	398	10 US-09-860-352-11099	Sequence 11099, A
15	20	1.0	370	13 US-10-027-632-140660	Sequence 140660, Sequence 334, App
16	20	1.0	11102	14 US-10-205-823-334	Sequence 334, App

17	20	1.0	180557	13 US-10-003-806-6	Sequence 6, Appl1
18	20	1.0	180557	13 US-10-003-806-9	Sequence 9, Appl1
19	19	0.9	341	10 US-09-867-701-8841	Sequence 8841, Ap
20	19	0.9	341	10 US-09-867-701-8888	Sequence 8888, Ap
21	19	0.9	349	9 US-09-759-143-649	Sequence 649, App
22	19	0.9	349	9 US-09-780-669-649	Sequence 649, App
23	19	0.9	349	9 US-09-822-827-649	Sequence 649, App
24	19	0.9	349	10 US-09-895-793-649	Sequence 649, App
25	19	0.9	349	10 US-09-895-814-649	Sequence 649, App
26	19	0.9	349	13 US-10-012-896-649	Sequence 649, App
27	19	0.9	405	10 US-09-867-701-9634	Sequence 9634, Ap
28	19	0.9	544	13 US-10-027-632-118925	Sequence 87347, A
29	19	0.9	786	13 US-10-027-632-118925	Sequence 118925, Sequence 147, App
30	19	0.9	2874	13 US-10-027-632-111785	Sequence 26, Appl1
31	19	0.9	3273	12 US-09-871-392-26	Sequence 52, Appl1
32	19	0.9	3274	13 US-10-002-600-52	Sequence 32, App
33	19	0.9	4224	9 US-09-815-242-9444	Sequence 32, App
34	19	0.9	4537	13 US-10-108-605-322	Sequence 2201, Ap
35	19	0.9	6546	12 US-10-017-161-2201	Sequence 147, App
36	19	0.9	8742	9 US-09-764-891-6217	Sequence 6217, Ap
37	19	0.9	10925	14 US-10-032-393-7237	Sequence 2257, Ap
38	19	0.9	62944	10 US-09-854-456-2237	Sequence 243, Appl1
39	19	0.9	75899	9 US-09-854-456-2237	Sequence 17017, A
40	19	0.9	684973	10 US-09-263-959-1	Sequence 8, Appl1
41	19	0.9	25	14 US-10-098-2638-17017	Sequence 765, App
42	18	0.9	47	9 US-09-816-825-8	
43	18	0.9	47	13 US-10-007-262-8	
44	18	0.9	129	9 US-09-759-143-765	
45	18	0.9			

ALIGNMENTS

RESULT 1
US-09-816-825-1
Sequence 1, Application US/09816825
Patent No. US20010051370A1
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107CON
CURRENT APPLICATION NUMBER: US/09/816, 825
CURRENT FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/045, 284
PRIOR FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2032
TYPE: DNA
ORGANISM: Homo sapiens
US-09-816-825-1

Query Match
Best Local Similarity 75.7%: Score 1547; DB 9; Length 2032;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAAAGCTTCCTCAGTTCAGCAACATGCTACCTTAAATGAAGCTCTGCTGTTCT 165
|||
Db 160 CAAAGCTTCCTCAGTTCAGCAACATGCTACCTTAAATGAAGCTCTGCTGTTCT 219
|||
QY 166 GGTTCCTCAGTTCAGTTCCTGCTATCTTCCATGTCAGCCACCAATCAGCTC 225
|||
Db 220 GGTTCCTCAGTTCAGTTCCTGCTATCTTCCATGTCAGCCACCAATCAGCTC 279
|||
QY 226 CCGTCTATGAGGACAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 285
|||
Db 280 CCGTCTATGAGGACAGCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 339
|||
QY 286 TGGCTCTCTCTTGTGGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 345

Db 340 TGGCTCTCTTTTGTGGGAGCTTTTGGGACAGCCAGATGTTTCTACCTGATGGA 399
 QY 346 GCCCGCTGGACAGTGTGATGATCTTCAAGACAGACCGCTGATGCTGCATGGC 405
 Db 400 GCCCGCTGGACAGTGTGATGATCTTCAAGACAGACCGCTGATGCTGCATGGC 459
 QY 406 TGTGGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTTGATGCTACAT 465
 Db 460 TGTGGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTTGATGCTACAT 519
 QY 466 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAAACAGCCGGCCCTGTG 525
 Db 520 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAAACAGCCGGCCCTGTG 579
 QY 526 TTCTGACCTGCTGTGACATCATCCACAAAGATGAATCATCCCGGGCTCACTGAC 585
 Db 580 TTCTGACCTGCTGTGACATCATCCACAAAGATGAATCATCCCGGGCTCACTGAC 639
 QY 586 GCTCTGTGATCAACAGCCCTTTGAGGTGTGAGAAAGCCCTGCCCTCTACAGCCA 645
 Db 640 GCTCTGTGATCAACAGCCCTTTGAGGTGTGAGAAAGCCCTGCCCTCTACAGCCA 699
 QY 646 CGTGGTCTCAAGAGAGTGGCTCTTCAACCTGACGTCCTTACCCCTGCTGAAGA 705
 Db 700 CGTGGTCTCAAGAGAGTGGCTCTTCAACCTGACGTCCTTACCCCTGCTGAAGA 759
 QY 706 CCCCTCCCTCAACCTGATCTGACAGTGTGCGGAGCCCGGGCCCTGTTCCGTTG 765
 Db 760 CCCCTCCCTCAACCTGATCTGACAGTGTGCGGAGCCCGGGCCCTGTTCCGTTG 819
 QY 766 CCGAGAACGACAAAGGAGATCTCATGTATGACAGTCCGATTTGTATGGGAGCATGA 825
 Db 820 CCGAGAACGACAAAGGAGATCTCATGTATGACAGTCCGATTTGTATGGGAGCATGA 879
 QY 826 GCAAAACCTCAAGAGAGAGACCAACCTACTATGATGATGACAGTATGCGCAAGCCA 885
 Db 880 GCAAAACCTCAAGAGAGAGACCAACCTACTATGATGATGACAGTATGCGCAAGCCA 939
 QY 886 GCTGGAGATCTACAGAACCATCCAGTCTTGTGCCAAGGCCCTGACAGAACGCTACCTGT 945
 Db 940 GCTGGAGATCTACAGAACCATCCAGTCTTGTGCCAAGGCCCTGACAGAACGCTACCTGT 999
 QY 946 TGTGGCTATGAG 1005
 Db 1000 TGTGGCTATGAG 1059
 QY 1006 CGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1065
 Db 1060 CGTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1119
 QY 1066 GGGCATGGGTGACACAGCTTTTCCACAAATGCGAAGGATGCCCTTAATGTCTCCAGGC 1125
 Db 1120 GGGCATGGGTGACACAGCTTTTCCACAAATGCGAAGGATGCCCTTAATGTCTCCAGGC 1179
 QY 1126 TTTGGGCTGGTCTTTTGGGATGAAAGGTTTCTGACATTCAGAAAGCCCTGTGGCATGC 1185
 Db 1180 TTTGGGCTGGTCTTTTGGGATGAAAGGTTTCTGACATTCAGAAAGCCCTGTGGCATGC 1239
 QY 1186 CATGAATTTGCTGGGCTACCGCAGCTGATGATGAAGAAGACAGAAAGCTGTGGCT 1245
 Db 1240 CATGAATTTGCTGGGCTACCGCAGCTGATGATGAAGAAGACAGAAAGCTGTGGCT 1299
 QY 1246 GGAATCTTGTCTACCTGAGACTGTCCCTGAGCAATTCACCTAAGAGGGTTGAGAAAGCTTT 1305
 Db 1300 GGAATCTTGTCTACCTGAGACTGTCCCTGAGCAATTCACCTAAGAGGGTTGAGAAAGCTTT 1359
 QY 1306 TGTGACACCTGCTGACAGCTGACAGCTGATGATGAAGAAGACAGAAAGCTGTGGCT 1365
 Db 1360 TGTGACACCTGCTGACAGCTGACAGCTGATGATGAAGAAGACAGAAAGCTGTGGCT 1419
 QY 1366 TGTCTAGGCTTACTACATGCTGTGGATCACACTGATGATGATGATGATGATGATGATGAT 1425

Db 1420 TGTCTAGGCTTACTACATGCTGTGGGATATCACACTGATGATGATGATGATGATGATGATGAT 1479
 QY 1426 TGTCTAGGCTTACTACATGCTGTGGGATATCACACTGATGATGATGATGATGATGATGATGATGAT 1485
 Db 1480 TGTCTAGGCTTACTACATGCTGTGGGATATCACACTGATGATGATGATGATGATGATGATGATGAT 1539
 QY 1486 TATGTAGACAGCAGATCCACAGAGTGAAGAGGATATGCTCTTCTTTCTTGTATCT 1545
 Db 1540 TATGTAGACAGCAGATCCACAGAGTGAAGAGGATATGCTCTTCTTTCTTGTATCT 1599
 QY 1546 TCTGTGTGGGAGAGCTTTCAGAGATCTTGTGGGCTGAGAGCCCTATTAAGACAGACAGT 1605
 Db 1600 TCTGTGTGGGAGAGCTTTCAGAGATCTTGTGGGCTGAGAGCCCTATTAAGACAGACAGT 1659
 QY 1606 ATCAGTGAATGATCATCAATCAACCTCCCTGTCACATCTTCCCAATGGGAGATGATCT 1665
 Db 1660 ATCAGTGAATGATCATCAATCAACCTCCCTGTCACATCTTCCCAATGGGAGATGATCT 1719
 QY 1666 TTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
 Db 1720 TTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1757
 RESULT 2
 US-10-007-262-2
 ; Sequence 2, Application US/10007262
 ; Publication No. US20020164748A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Blstrup, Annette
 ; APPLICANT: Rosen, Steven D.
 ; APPLICANT: Tangemann, Kirsten
 ; APPLICANT: Hemmerlich, Stefan
 ; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
 ; FILE REFERENCE: 6510-107CIP
 ; CURRENT APPLICATION NUMBER: US/10/007,262
 ; PRIOR FILING DATE: 2001-11-08
 ; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
 ; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
 ; NUMBER OF SEQ ID NOS: 8
 ; SOFTWARE: FastSeq for Windows Version 3.0
 ; SEQ ID NO 2
 ; LENGTH: 2032
 ; TYPE: DNA
 ; ORGANISM: H. sapiens
 ; US-10-007-262-2
 Query Match 75.7%; Score 1547; DB 13; Length 2032;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
 QY 106 CAAGGTCTTCACCTTACAGCAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTTCT 165
 Db 160 CAAGGTCTTCACCTTACAGCAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTTCT 219
 QY 166 GCTTTCAGATGGCCATCTTGGCTTATCTTCCACATGTATACGCCAACAATCAGCTC 225
 Db 220 GCTTTCAGATGGCCATCTTGGCTTATCTTCCACATGTATACGCCAACAATCAGCTC 279
 QY 226 CCTGTATGAAGGACAGCCCGAGGAGATGACAGTGTGCTGCTTCTTCCGTGGGCTC 285
 Db 280 CCTGTATGAAGGACAGCCCGAGGAGATGACAGTGTGCTGCTTCTTCCGTGGGCTC 339
 QY 286 TGGCTCTCTTTTGTGGGAGAGCTTTTGGGAGACCCAGATGTTTCTACCTGATGGA 345
 Db 340 TGGCTCTCTTTTGTGGGAGAGCTTTTGGGAGACCCAGATGTTTCTACCTGATGGA 399
 QY 346 GCCCGCTGGACAGTGTGATGATCTTCAAGACAGACACCGCTGATGCTGCATGGC 405
 Db 400 GCCCGCTGGACAGTGTGATGATCTTCAAGACAGACACCGCTGATGCTGCATGGC 459
 QY 406 TGTGGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTTGATGCTACAT 465
 Db 460 TGTGGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTTGATGCTACAT 519

466 GGAACCTGTCCTCCCGGAGACAGTCACCTCTTTCAGTGGGAGAACAGCCGGCCCTGTG 525
520 GGAACCTGTCCTCCCGGAGAGAGTCACCTCTTTCAGTGGGAGAACAGCCGGCCCTGTG 579
526 TTCTGACACCTGCTGTGATCATCCCAAGATGAATATCCCCGGGCTCACTGCAG 585
580 TTCTGACACCTGCTGTGATCATCCCAAGATGAATATCCCCGGGCTCACTGCAG 639
586 GCTCTGTGAGTCAACAGCCCTTGAAGTGGGAGAGGCGTGGCCCTCTACAGCA 645
640 GCTCTGTGAGTCAACAGCCCTTGAAGTGGGAGAGGCGTGGCCCTCTACAGCA 699
646 GCTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAAGA 705
700 GCTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAAGA 759
706 CCCCCTCAACCTGATATGCTGACACCTGTCCGGAGCCCGGGCCGTTCCTGTC 765
760 CCCCCTCAACCTGATATGCTGACACCTGTCCGGAGCCCGGGCCGTTCCTGTC 819
766 CCGAGAACGCAAGAGAGATCATGATTGACAGTGGATGGGCGAGCATGA 825
820 CCGAGAACGCAAGAGAGATCATGATTGACAGTGGATGGGCGAGCATGA 879
826 GCAAAAACCTCAAGAGAGAGACCAACCTTATGATGATGAGGTCACTGTGCAAAAGCA 885
880 GCAAAAACCTCAAGAGAGAGACCAACCTTATGATGATGAGGTCACTGTGCAAAAGCA 939
886 GCTGAGATCTACAAAGACATCCAGTCTTCCAGAGCCCTGCAAGAGACGATCTGCT 945
940 GCTGAGATCTACAAAGACATCCAGTCTTCCAGAGCCCTGCAAGAGACGATCTGCT 999
946 TGTGCGTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTCCGAATGTGATTT 1005
1000 TGTGCGTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTCCGAATGTGATTT 1059
1006 GGTGAGATGGAATTTCTTGGCCCATCTTTCAGACCTGGGTGATTAATCAACCCGAGCAA 1065
1060 GGTGAGATGGAATTTCTTGGCCCATCTTTCAGACCTGGGTGATTAATCAACCCGAGCAA 1119
1066 GGTGAGATGGAATTTCTTGGCCCATCTTTCAGACCTGGGTGATTAATCAACCCGAGCAA 1125
1120 GGTGAGATGGAATTTCTTGGCCCATCTTTCAGACCTGGGTGATTAATCAACCCGAGCAA 1179
1126 TTTGGCGTGTGCTTGGCCCATATGAAGGTTTCTGACTTCAGAAACCTGTGGCGATGC 1185
1180 TTTGGCGTGTGCTTGGCCCATATGAAGGTTTCTGACTTCAGAAACCTGTGGCGATGC 1239
1186 CATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAACAAGAACAGAAACCTGTGCT 1245
1240 CATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAACAAGAACAGAAACCTGTGCT 1299
1246 GATCTTCTGTCTACCTGAGCTGCTCTGAGCAATCACTACAGAGGTTGAGAGGCTT 1305
1300 GATCTTCTGTCTACCTGAGCTGCTCTGAGCAATCACTACAGAGGTTGAGAGGCTT 1359
1306 TGTGCGACCTGTGAGAGCTGCTGATGATTTCTGATGATTTCTGAGCTTGTGCTTACA 1365
1360 TGTGCGACCTGTGAGAGCTGCTGATGATTTCTGATGATTTCTGAGCTTGTGCTTACA 1419
1366 TGTGAGAGCTTAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1425
1420 TGTGAGAGCTTAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1479
1426 TGTGAGAGCTTAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1485
1480 TGTGAGAGCTTAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1539
1486 TGTGAGAGCTTAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1545
1540 TGTGAGAGCTTAACTATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1599

QY 1546 TCTGTGTGGGAGACTTTCAGAGACTTGTGGCTTGAGAGGCTTAAAGCAGACAGT 1605
DB 1600 TCTGTGTGGGAGACTTTCAGAGACTTGTGGCTTGAGAGGCTTAAAGCAGACAGT 1659
QY 1606 ATCACTGATTTGATTCATTAACCTCCCTGTCCATCTTGTCCCAATGGGAATGATCT 1665
DB 1660 ATCACTGATTTGATTCATTAACCTCCCTGTCCATCTTGTCCCAATGGGAATGATCT 1719
QY 1666 TTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
DB 1720 TTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1757

RESULT 3

US-09-998-2595
Sequence 2595, Application US/09989598
Patent No. US20020150922A1
GENERAL INFORMATION:
APPLICANT: Stolk, John A.
APPLICANT: Xu, Jianshun
APPLICANT: Chenault, Ruth A.
APPLICANT: Mesgher, Madelein Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
FILE REFERENCE: 210121.561
CURRENT FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 2606
SOFTWARE: Cortix Invention Disclosure Database
SEQ ID NO 2595
LENGTH: 505
TYPE: DNA
ORGANISM: Homo sapiens
US-09-998-598-2595

Query Match 19.7%; Score 403; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 8.5e-203;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGTCTGCTGGCAGCTGTGTCACCTCAGTCACTTTCTGTGATGCTTGTGAGCCTTGC 1360
DB 1 GGTCTGCTGGCAGCTGTGTCACCTCAGTCACTTTCTGTGATGCTTGTGAGCCTTGC 60
QY 1361 CTACATCTCTAGGCTTAACTATGATGATGATGATGATGATGATGATGATGATGATG 1420
DB 61 CTACATCTCTAGGCTTAACTATGATGATGATGATGATGATGATGATGATGATGATG 120
QY 1421 ACAGTGTCTCAAGCAGAGAGACTTTGTGTCATGCTTGTGTGTAAGAAACAGACTGGG 1480
DB 121 ACAGTGTCTCAAGCAGAGAGACTTTGTGTCATGCTTGTGTGTAAGAAACAGACTGGG 180
QY 1481 AACCTTATGTGAGAGACATCCACAGATGAAGAGGATTTCTTCTTCTTCTTCTTCTT 1540
DB 181 AACCTTATGTGAGAGACATCCACAGATGAAGAGGATTTCTTCTTCTTCTTCTTCTT 240
QY 1541 GATCTTCTGTCTGGGAGACTTTCAGAGACTTGTGGCGTGGAGGCTTAAAGCAGAC 1600
DB 241 GATCTTCTGTCTGGGAGACTTTCAGAGACTTGTGGCGTGGAGGCTTAAAGCAGAC 300
QY 1601 ACAGTATCAGTGAATGATCAATAACTCCCTGTCCATCTTGTGCCAATGGGAGATG 1660
DB 301 ACAGTATCAGTGAATGATCAATAACTCCCTGTCCATCTTGTGCCAATGGGAGATG 360
QY 1661 GATCTTTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
DB 361 GATCTTTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 403

RESULT 4

US-09-919-580-194
Sequence 194, Application US/09919580
Patent No. US20020110832A1
GENERAL INFORMATION:

APPLICANT: Pyle, Ruth
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.552
CURRENT APPLICATION NUMBER: US/09/919,580
CURRENT FILING DATE: 2001-07-30
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 194
LENGTH: 517
TYPE: DNA
ORGANISM: Homo sapiens
US-09-919-580-194

Query Match 19.7%; Score 403; DB 10; Length 517;
Best Local Similarity 100.0%; Pred. No. 8.5e-203; Indels 0; Gaps 0;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCTTGGCCACCTGCTGAGCCTGACATCTTCTCTGAATGCTTCTGAGCCTTGC 1360
DB 13 GGCTTGGCCACCTGCTGAGCCTGACATCTTCTCTGAATGCTTCTGAGCCTTGC 72
QY 1361 CTACATCTGAGCCTTACTACTGCTGCTGGGATTCACACTGAGTGTGCTGTC 1420
DB 73 CTACATCTGAGCCTTACTACTGCTGCTGGGATTCACACTGAGTGTGCTGTC 132
QY 1421 ACACGTGCTCAGACAGAGAGACTTGTCTCCATGCTTGTCTAGAAAACAGACGCGG 1480
DB 133 ACACGTGCTCAGACAGAGAGACTTGTCTCCATGCTTGTCTAGAAAACAGACGCGG 192
QY 1481 AACCTATGTGAGACACACATCCACAGTGAACAGGGATGCTCTTCTTCTTCTT 1540
DB 193 AACCTATGTGAGACACACATCCACAGTGAACAGGGATGCTCTTCTTCTTCTT 252
QY 1541 GATCTTCTGCTGCTGAGCCTGACATCTTGTGCTGGGAGGCTTCTTGAAGCAGAC 1600
DB 253 GATCTTCTGCTGCTGAGCCTGACATCTTGTGCTGGGAGGCTTCTTGAAGCAGAC 312
QY 1601 ACAGTATGAGTGAATGATCCATAAACCTCCCTGTCACATCTTGGCCCAATGGGAGAG 1660
DB 313 ACAGTATGAGTGAATGATCCATAAACCTCCCTGTCACATCTTGGCCCAATGGGAGAG 372
QY 1661 GATCTTCTGAGAGCTCACCAGATTTTCCACAGAGATGC 1703
DB 373 GATCTTCTGAGAGCTCACCAGATTTTCCACAGAGATGC 415

RESULT 5

US-09-919-580-264
Sequence 264, Application US/09919580
Patent No. US20020110832A1
GENERAL INFORMATION:
APPLICANT: Pyle, Ruth
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.552
CURRENT APPLICATION NUMBER: US/09/919,580
CURRENT FILING DATE: 2001-07-30
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 264
LENGTH: 389
TYPE: DNA
ORGANISM: Homo sapiens
US-09-919-580-264

Query Match 5.6%; Score 114; DB 10; Length 389;
Best Local Similarity 100.0%; Pred. No. 1.3e-49;
Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGCCATCTGTATTAATAAATCCCAATTAAGTCTGTGTAATGTCCTTTATG 1989
DB 266 TGCCATCTGTATTAATAAATCCCAATTAAGTCTGTGTAATGTCCTTTATG 325
QY 1990 CTCTTAATTAATTAAGAGTAATGTCATTTTATGAGTCCCTAAAAA 2043
DB 326 CTCTTAATTAATTAAGAGTAATGTCATTTTATGAGTCCCTAAAAA 379

RESULT 6

US-09-927-602-1
Sequence 1, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
TITLE OF INVENTION: Dystrophy
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2544
TYPE: DNA
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: CDS
LOCATION: (693)...(1877)
US-09-927-602-1

Query Match 2.9%; Score 59; DB 9; Length 2544;
Best Local Similarity 100.0%; Pred. No. 1.9e-20;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGGCTCTTACAGCAGGCTGCTCAAGAGGTGCGCTTCTTCACTGCAG 682
DB 1180 AGGCTGCGGCTCTTACAGCAGGCTGCTCAAGAGGTGCGCTTCTTCACTGCAG 1238
RESULT 7
US-09-927-602-38
Sequence 38, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
TITLE OF INVENTION: Dystrophy
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 38
LENGTH: 48436
TYPE: DNA
ORGANISM: Homo Sapien
US-09-927-602-38

Query Match 2.9%; Score 59; DB 9; Length 48436;
Best Local Similarity 100.0%; Pred. No. 1.9e-20;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGGCTCTTACAGCAGGCTGCTCAAGAGGTGCGCTTCTTCACTGCAG 682
DB 48024 AGGCTGCGGCTCTTACAGCAGGCTGCTCAAGAGGTGCGCTTCTTCACTGCAG 48082

RESULT 8
US-09-816-825-7
; Sequence 7, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816,825
; PRIOR FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-816-825-7

Query Match
Best Local Similarity 1.8%; Score 37; DB 9; Length 37;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 830 AACTCAAGAAGAGGACCAACCTACTATGTGATGC 866
Db 1 AACTCAAGAAGAGGACCAACCTACTATGTGATGC 37

RESULT 9
US-10-007-262-7
; Sequence 7, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007,262
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-007-262-7

Query Match
Best Local Similarity 1.8%; Score 37; DB 13; Length 37;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 830 AACTCAAGAAGAGGACCAACCTACTATGTGATGC 866
Db 1 AACTCAAGAAGAGGACCAACCTACTATGTGATGC 37

RESULT 10
US-09-867-701-9458
; Sequence 9458, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9458
; LENGTH: 167
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9458

Query Match
Best Local Similarity 1.1%; Score 22; DB 10; Length 167;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 GAAGACGACAGAGGCTAGAGG 44
Db 3 GAAGACGACAGAGGCTAGAGG 24

RESULT 11
US-09-867-701-9067
; Sequence 9067, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9067
; LENGTH: 171
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9067

Query Match
Best Local Similarity 1.1%; Score 22; DB 10; Length 171;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 GAAGACGACAGAGGCTAGAGG 44
Db 9 GAAGACGACAGAGGCTAGAGG 30

RESULT 12
US-09-867-701-9185
; Sequence 9185, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; TITLE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9185
; LENGTH: 187
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9185

Query Match 1.1%; Score 22; DB 10; Length 187;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACGACGAAGGCTAGAG 44
DB 32 GAAGACGACGAAGGCTAGAG 53

RESULT 13

US-09-867-701-9284
Sequence 9284, Application US/09867701
Patent No. US2002013237A1
GENERAL INFORMATION:
APPLICANT: Aglate, Paul A.
APPLICANT: Jones, Robert
APPLICANT: Harlocker, Susan L.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.497
CURRENT APPLICATION NUMBER: US/09/867,701
CURRENT FILING DATE: 2001-05-29
NUMBER OF SEQ ID NOS: 10912
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 9284
LENGTH: 191
TYPE: DNA
ORGANISM: Homo sapien
US-09-867-701-9284

Query Match 1.1%; Score 22; DB 10; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACGACGAAGGCTAGAG 44
DB 29 GAAGACGACGAAGGCTAGAG 50

RESULT 14

US-09-960-352-11099/C
Sequence 11099, Application US/09960352
Patent No. US20020137139A1
GENERAL INFORMATION:
APPLICANT: Warren, Wesley C.
APPLICANT: Tao, Nengbing
APPLICANT: Byatt, John C.
APPLICANT: Mathalagan, Nagappan
TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
FILE REFERENCE: 16511.006/37-21(10298)C
CURRENT APPLICATION NUMBER: US/09/960,352
CURRENT FILING DATE: 2001-09-24
NUMBER OF SEQ ID NOS: 15112
SEQ ID NO 11099
LENGTH: 398
TYPE: DNA
ORGANISM: Bos taurus
OTHER INFORMATION: Clone ID: 48-BOVMS1-008-Q1-E1-D8
US-09-960-352-11099

Query Match 1.0%; Score 20; DB 10; Length 398;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GATGGCCATCTGGCTCTAT 194
DB 240 GATGGCCATCTGGCTCTAT 221

RESULT 15

US-10-027-632-140660
Sequence 140660, Application US/10027632

GENERAL INFORMATION:
APPLICANT: Wang, David G.
TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
TITLE OF INVENTION: Polymorphisms in the Human Genome
FILE REFERENCE: 108827.129
CURRENT APPLICATION NUMBER: US/10/027,632
CURRENT FILING DATE: 2002-04-30
PRIOR APPLICATION NUMBER: US 60/218,006
PRIOR FILING DATE: 2000-07-12
PRIOR APPLICATION NUMBER: US 60/198,676
PRIOR FILING DATE: 2000-04-20
PRIOR APPLICATION NUMBER: US 60/193,483
PRIOR FILING DATE: 2000-03-29
PRIOR APPLICATION NUMBER: US 60/185,218
PRIOR FILING DATE: 2000-02-24
PRIOR APPLICATION NUMBER: US 60/167,363
PRIOR FILING DATE: 1999-11-23
PRIOR APPLICATION NUMBER: US 60/156,358
PRIOR FILING DATE: 1999-09-28
PRIOR APPLICATION NUMBER: US 60/146,002
PRIOR FILING DATE: 1999-08-09
NUMBER OF SEQ ID NOS: 325720
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 140660
LENGTH: 570
TYPE: DNA
ORGANISM: Human
US-10-027-632-140660

Query Match 1.0%; Score 20; DB 13; Length 570;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 GAAAGACCCCTCCCTCAACC 719
DB 273 GAAAGACCCCTCCCTCAACC 292

Search completed: August 15, 2003, 00:56:03
Job time: 538 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 23:39:27 ; Search time 7280 Seconds

(without alignments)
11480.545 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattccattgtgttgta.....tgggaccccaaaaaaaaaa 2043

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Word size: 0

Search: 2888711 seqs, 2045481386 residues

Minimum DB seq length: 50

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

GenBank:*

1: gb_ba:*

2: gb_htg:*

3: gb_in:*

4: gb_om:*

5: gb_ov:*

6: gb_pat:*

7: gb_ph:*

8: gb_pl:*

9: gb_pr:*

10: gb_ro:*

11: gb_sts:*

12: gb_sy:*

13: gb_un:*

14: gb_vl:*

15: em_ba:*

16: em_fun:*

17: em_hum:*

18: em_in:*

19: em_mu:*

20: em_om:*

21: em_or:*

22: em_ov:*

23: em_pat:*

24: em_ph:*

25: em_pl:*

26: em_ro:*

27: em_sts:*

28: em_un:*

29: em_vl:*

30: em_htg_hum:*

31: em_htg_inv:*

32: em_htg_other:*

33: em_htg_mus:*

34: em_htg_pln:*

35: em_htg_rod:*

36: em_htg_man:*

37: em_htg_vrl:*

38: em_sy:*

39: em_htg_hum:*

40: em_htg_mus:*

41: em_htg_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1722	84.3	1992	9	AF280088
2	1633	80.9	2037	9	BC035282
3	1598	78.2	2011	9	AK026635
4	1597	78.2	183228	9	AC010547
5	1597	78.2	206943	2	AC138848
6	1547	75.7	1979	6	BD127258
7	1547	75.7	1979	6	AK074746
8	1547	75.7	2032	6	AR203335
9	1547	75.7	2032	6	BD134772
10	1547	75.7	2032	9	AF131235
11	1262	61.8	1333	9	AF149783
12	1153	56.4	2065	6	AR310485
13	695	34.0	877	6	BD124832
14	695	34.0	877	6	BD126950
15	403	19.7	517	6	AX381256
16	114	5.6	389	6	AX381326
17	105	5.1	548	6	BD125617
18	59	2.9	1462	9	AF176839
19	59	2.9	1647	6	AX327330
20	59	2.9	2170	9	AF176838
21	59	2.9	2544	9	AF219990
22	59	2.9	3278	9	AF246718
23	59	2.9	3786	9	AF280086
24	59	2.9	71503	9	AF219991
25	59	2.9	157358	9	AC009163
26	59	2.9	194832	9	AC025287
27	59	2.9	208185	2	AC009105
28	29	1.3	1926	6	BD134773
29	27	1.3	2201	10	AF131236
30	31	1.3	22745	2	AC130174
31	27	1.3	232745	2	AC130174
32	33	1.2	228569	2	AC121740
33	33	1.2	235173	2	AC133449
34	33	1.2	240901	2	AC096328
35	33	1.2	242253	2	AC095651
36	37	1.2	246877	2	AC137365
37	38	1.2	258097	2	AC129764
38	39	1.2	271230	2	AC094919
39	40	1.1	1740	10	AF176841
40	41	1.1	1989	10	AF176840
41	42	1.1	110000	2	AC099262-1
42	23	1.1	159170	2	AC115914
43	23	1.1	192365	2	AC139891
44	23	1.1	194439	2	AC109022
45	23	1.1	194439	2	AC109022

ALIGNMENTS

RESULT 1

AF280088

LOCUS

DEFINITION Homo sapiens l-selectin ligand sulfotransferase GST-3 mRNA,

ACCESSION AF280088

VERSION AF280088.1

KEYWORDS

SOURCE

ORGANISM

Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 1992)

Hemmerich, S., Lee, J.K., Bhakta, S., Bistrup, A., Ruddle, N.R. and Rosen, S.D.


```

OY 951 GGTATGAGAGAGCTGGCTGAGCCCTGTGGCCAGACTTCCGAAATGATGATTCGTGG 1010
DB 901 GGTATGAGAGAGCTGGCTGAGCCCTGTGGCCAGACTTCCGAAATGATGATTCGTGG 960
OY 1011 GATTGGAATTTCTTGGCCCATTTTCAGACTGGGTGATTAACATCACCCGAGCAGGCGCA 1070
DB 961 GATTGGAATTTCTTGGCCCATTTTCAGACTGGGTGATTAACATCACCCGAGCAGGCGCA 1020
OY 1071 TGGGTGAGACGAGCTTTCACAAATGCGAGGATGCGCTTATGTCCTCCAGGCTTGGC 1130
DB 1021 TGGGTGAGACGAGCTTTCACAAATGCGAGGATGCGCTTATGTCCTCCAGGCTTGGC 1080
OY 1131 GCTGCTCTTGGCCCATTAATGAAGTTTCTGCACTTGAAGACCTGTGGCATGCGATGA 1190
DB 1081 GCTGCTCTTGGCCCATTAATGAAGTTTCTGCACTTGAAGACCTGTGGCATGCGATGA 1140
OY 1191 ATTGCTGGGCTACGCGCCAGCTGATGTAACAGAAAGAAACCTGTGGCATGCGATGA 1250
DB 1141 ATTGCTGGGCTACGCGCCAGCTGATGTAACAGAAAGAAACCTGTGGCATGCGATGA 1200
OY 1251 TTCTGTCTACCTGAGACTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGGCTTGTCTG 1310
DB 1201 TTCTGTCTACCTGAGACTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGGCTTGTCTG 1260
OY 1311 CCACCTGGGTGAGGCTCAGTACTTCTCTGAAATGCTTCTGAGCCTTGCCTACATCTCT 1370
DB 1261 CCACCTGGGTGAGGCTCAGTACTTCTCTGAAATGCTTCTGAGCCTTGCCTACATCTCT 1320
OY 1371 GAGCCTTAACATGATGTCTGTGGGTATCAGACTGATGATGATGTTGTCACACGCTGCTC 1430
DB 1321 GAGCCTTAACATGATGTCTGTGGGTATCAGACTGATGATGATGTTGTCACACGCTGCTC 1380
OY 1431 AAGCAGAAAGAGCTTTTGTGTCATGCTTGTCTAGAAAACAGACTGGGGAACCTTATGT 1490
DB 1381 AAGCAGAAAGAGCTTTTGTGTCATGCTTGTCTAGAAAACAGACTGGGGAACCTTATGT 1440
OY 1491 GAGCAGCAATCCCAACCAATGAAACAGGATTTGCTCTCTCTTCTTCTGATCTTCTG 1550
DB 1441 GAGCAGCAATCCCAACCAATGAAACAGGATTTGCTCTCTCTTCTTCTGATCTTCTG 1500
OY 1551 TCTGGGCAAGCTTTCAGAGACTTGTGGGCTTGAGAGGCTTATTAAGCAGCAGAGATTCAG 1610
DB 1501 TCTGGGCAAGCTTTCAGAGACTTGTGGGCTTGAGAGGCTTATTAAGCAGCAGAGATTCAG 1560
OY 1611 TGGATTTGATCATTAACCTCCCTGTCCATCTTTCGCCAATGAGGGAATGATCTTTCAC 1670
DB 1561 TGGATTTGATCATTAACCTCCCTGTCCATCTTTCGCCAATGAGGGAATGATCTTTCAC 1620
OY 1671 CAAGAGCTCAGCAGATTTTCCACAGAGATGC 1703
DB 1621 CAAGAGCTCAGCAGATTTTCCACAGAGATGC 1653

```

```

RESULT 3
AK026635
LOCUS 2011 bp mRNA linear prt 29-sep-2000
DEFINITION Homo sapiens cDNA: FLJ22982 fis, clone KAT11454, highly similar to
AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase
mRNA.
ACCESSION AK026635
VERSION AK026635.1 GI:10439531
KEYWORDS oligo cloning; fis (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (sites)
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
AUTHORS Matenabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H.,
Ota, T., Suzuki, Y., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Nakamura, Y., Isogai, T. and Sugano, S.
NEBO human cDNA sequencing project
Unpublished
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 2011)

```

```

AUTHORS Sugano, S., Suzuki, Y., Ota, T., Odayashi, M., Nishi, T., Isogai, T.,
Shibahara, T., Tanaka, T. and Nakamura, Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdna@leims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NEBO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction, 5'- & 3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source location/Qualifiers
1..2011
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="KAT11454"
/cell_line="KATO III"
/cell_type="signet-ring cell carcinoma"
/clone_1lb="KAT"
/notes="cloning vector pME18SFL3"
misc_feature 1..2011
/notes="highly similar to AF131235 Homo sapiens
N-acetylglucosamine 6-O-sulfotransferase mRNA"
BASE COUNT 482 a 541 c 475 g 513 t
ORIGIN
Query Match 78.2%; Score 1598; DB 9; Length 2011;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1598; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 106 CAAGGCTCCAGTTCAGCAATGCTACTGCTAAAGAAAGAGCTGCTCTGTTCT 165
DB 50 CAAGGCTCCAGTTCAGCAATGCTACTGCTAAAGAAAGAGCTGCTCTGTTCT 109
OY 166 GGTTCCTCCAGATGGCCATCTTGGCTCTATCTTCTCCAGATGTCAGCAGCAATCAGCTC 225
DB 110 GGTTCCTCCAGATGGCCATCTTGGCTCTATCTTCTCCAGATGTCAGCAGCAATCAGCTC 169
OY 226 CCTGTATGAAGGACACACCCGAGCGCATGTCAGCTGCTGCTCTCTCTGCGCTC 285
DB 170 CCTGTATGAAGGACACACCCGAGCGCATGTCAGCTGCTGCTCTCTCTGCGCTC 229
OY 286 TGGCTCTCTTGTGTGGGAGAGCTTTTGGGAGAGACCCAGATGTTTCTACCTGATGA 345
DB 230 TGGCTCTCTTGTGTGGGAGAGCTTTTGGGAGAGACCCAGATGTTTCTACCTGATGA 289
OY 346 GCCCGCTGGCAGCTGTGATGACCTTTCAGCAGAGACACCGCTGATGCTGACATGCG 405
DB 290 GCCCGCTGGCAGCTGTGATGACCTTTCAGCAGAGACACCGCTGATGCTGACATGCG 349
OY 406 TGTGCGGATCTGATACGCGCGCTCTTCTGTGCGCAGATGAGCGCTTTGATGCTCACT 465
DB 350 TGTGCGGATCTGATACGCGCGCTCTTCTGTGCGCAGATGAGCGCTTTGATGCTCACT 409
OY 466 GGAACCTGTGTCGCCGAGAGACAGTCCAGCTCTTTCAGTGGGAGAAACAGCGGCGCTG 525
DB 410 GGAACCTGTGTCGCCGAGAGACAGTCCAGCTCTTTCAGTGGGAGAAACAGCGGCGCTG 469
OY 526 TTCTGCACTGCTGTGATCATATCCACAAAGATGAATCATCCCGGCTCACTGAC 585
DB 470 TTCTGCACTGCTGTGATCATATCCACAAAGATGAATCATCCCGGCTCACTGAC 529
OY 586 GCTCTGTGCACTCAACAGCCCTTTGAGAGTGTGGAAGAGCGCTGCTCTTACAGCA 645
DB 530 GCTCTGTGCACTCAACAGCCCTTTGAGAGTGTGGAAGAGCGCTGCTCTTACAGCA 589
OY 646 CGTGTGCTCAAGAGAGGTGCTCTTTCACCTGACGTCCTCCGCTGCTGAAGA 705

```

Db 590 CGTGTGCTCAGAGAGTGGCTTCTTCAACCTGACCTGCCCTCTACCCGCTGTGANA 649
 QY 706 CCCCTCCCTACCTGCATATCGTGCACGTGTCGGGACCCCGGGCGGTTCGCTTC 765
 Db 650 CCCCTCCCTACCTGCATATCGTGCACGTGTCGGGACCCCGGGCGGTTCGCTTC 709
 QY 766 CCGAGAACGACAAAGGAGATCTCATGATTTGACAGTGCATTTGTGTGGGGAGCATGA 825
 Db 710 CCGAGAACGACAAAGGAGATCTCATGATTTGACAGTGCATTTGTGTGGGGAGCATGA 769
 QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTCTATGTATGAGTCTATCTGCCAAACCA 885
 Db 770 GCAAAAACCTCAAGAGAGAGACCAACCTCTATGTATGAGTCTATCTGCCAAACCA 829
 QY 886 GCTGAGATCTACAGACATCCAGTCTCTGCCCAAGGCCCTGAGAGAGGCTACCTGCT 945
 Db 830 GCTGAGATCTACAGACATCCAGTCTCTGCCCAAGGCCCTGAGAGAGGCTACCTGCT 889
 QY 946 TGTGCGCTATGAGAGACCTGGCTGAGGCCCTGTGGCCAGACTTCCGAATGTATGAAT 1005
 Db 890 TGTGCGCTATGAGAGACCTGGCTGAGGCCCTGTGGCCAGACTTCCGAATGTATGAAT 949
 QY 1006 CGTGGGATTTGGAATTTCTGGCCCATCTTCAGACCTGGGTGCATTAACCCGAGCAA 1065
 Db 950 CGTGGGATTTGGAATTTCTGGCCCATCTTCAGACCTGGGTGCATTAACCCGAGCAA 1009
 QY 1066 GGGCATGGTGCACACGCTTTCACACAAATGCGAGGATGCCCTTAATCTGCCAGGC 1125
 Db 1010 GGGCATGGTGCACACGCTTTCACACAAATGCGAGGATGCCCTTAATCTGCCAGGC 1069
 QY 1126 TTGGCGCTGCTTTTGGCCCATATGAAAAGTTTCTCGACTTCAGAAACCTGTGGCGATGC 1185
 Db 1070 TTGGCGCTGCTTTTGGCCCATATGAAAAGTTTCTCGACTTCAGAAACCTGTGGCGATGC 1129
 QY 1186 CATGAATTTCTGTGGCTACCCGCACTGCAATCTGAACAGAAACCACTGTTGCT 1245
 Db 1130 CATGAATTTCTGTGGCTACCCGCACTGCAATCTGAACAGAAACCACTGTTGCT 1189
 QY 1246 GGATCTTCTGTACCTGAGCTGCTGAGCAAAATCCACTAAGAGGCTGAGAAGCTT 1305
 Db 1190 GGATCTTCTGTACCTGAGCTGCTGAGCAAAATCCACTAAGAGGCTGAGAAGCTT 1249
 QY 1306 TGTGCGACCTGCTGTCAGCTGACCTGAGCAAAATCCACTAAGAGGCTGAGAAGCTT 1365
 Db 1250 TGTGCGACCTGCTGTCAGCTGACCTGAGCAAAATCCACTAAGAGGCTGAGAAGCTT 1309
 QY 1366 TCTCTGAGCTTAACATGATGTGTGGTATACACTGATGTGAGTGTGTCACAGC 1425
 Db 1310 TCTCTGAGCTTAACATGATGTGTGGTATACACTGATGTGAGTGTGTCACAGC 1369
 QY 1426 TGTCAAGCAGAGAGACTTTGTGTCATGCTGTCTGTAAGAAACAGATGGGGAACCT 1485
 Db 1370 TGTCAAGCAGAGAGACTTTGTGTCATGCTGTCTGTAAGAAACAGATGGGGAACCT 1429
 QY 1486 TATGTGAGCAGACATCCACACAGTGAACAGGGTATTTGCTTCTTCTTCTGATCT 1545
 Db 1430 TATGTGAGCAGACATCCACACAGTGAACAGGGTATTTGCTTCTTCTTCTGATCT 1489
 QY 1546 TCTGTCTGGGAGACTTCAAGAGCTTTGGCCCTGAGGCTTATTAAGACAGACAGT 1605
 Db 1490 TCTGTCTGGGAGACTTCAAGAGCTTTGGCCCTGAGGCTTATTAAGACAGACAGT 1549
 QY 1606 ATAGAGGAATGATCATTAACCTCCCTGCTTCCACATCTTGCCCAATGGGGAATGATCT 1665
 Db 1550 ATAGAGGAATGATCATTAACCTCCCTGCTTCCACATCTTGCCCAATGGGGAATGATCT 1609
 QY 1666 TTCACCAAGAGCTCAGACAGATTTTCCACAGAGATGC 1703
 Db 1610 TTCACCAAGAGCTCAGACAGATTTTCCACAGAGATGC 1647

RESULT 4
 AC010547

LOCUS AC010547 183228 bp DNA linear PRI 29-SEP-2001
 DEFINITION Homo sapiens chromosome 16 clone RP11-510M2, complete sequence.
 AC010547
 VERSION AC010547.9 GI:15808510
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 183228)
 AUTHORS DOE Joint Genome Institute.
 TITLE Sequencing of Human Chromosome 16
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 183228)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE 3 (bases 1 to 183228)
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (29-SEP-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT On Sep 29, 2001 this sequence version replaced gi:1589436.
 Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

 FEATURES
 source location/Qualifiers
 1..183228
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-510M2"
 BASE COUNT 51901 a 42098 c 40026 g 49203 t
 ORIGIN
 Query Match 78.2%; Score 1597; DB 9; Length 183228;
 Best local Similarity 100.0%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 107 AAGGCTTCACCTCAGCACAAATGCTACTGCTTAATAAATGAAGCTCTGCTGTTCTG 166
 QY 167 GTTTCACAGATGGGCACTTGGCTCTATTTCTTCACATGTACAGCCCAACATCAGCTCC 226
 Db 142644 GTTTCACAGATGGGCACTTGGCTCTATTTCTTCACATGTACAGCCCAACATCAGCTCC 142703
 QY 227 CTGCTATGAGAGCAGCCGAGCGCATCAGCTGCTGTTGCTTCTTCTGCGCTCT 286
 Db 142704 CTGCTATGAGAGCAGCCGAGCGCATCAGCTGCTGTTGCTTCTTCTGCGCTCT 142763
 QY 287 GGCTCTTCTTTGTGGGAGACTTTTGGGAGACCCAGACAGTCTTCTACCTGAGAG 346
 Db 142764 GGCTCTTCTTTGTGGGAGACTTTTGGGAGACCCAGACAGTCTTCTACCTGAGAG 142823

```

OY 347 CCGGCTGGACAGTGTGATGACCTTCAAGCAGACACCGCTGATGCTGACATGCGT 406
DB 142824 CCGGCTGGACAGTGTGATGACCTTCAAGCAGACACCGCTGATGCTGACATGCGT 142883
OY 407 GTCGGGATGTATACGGGGCTCTTCTTGTGACATGAGCGCTTGTATGCTTACATG 466
DB 142884 GTCGGGATGTATACGGGGCTCTTCTTGTGACATGAGCGCTTGTATGCTTACATG 142943
OY 467 GAACTGGTCCCGGAGACATGACAGCTTTCAGTGGAGAACAGCGGGCGCTGTGT 526
DB 142944 GAACTGGTCCCGGAGACATGACAGCTTTCAGTGGAGAACAGCGGGCGCTGTGT 143003
OY 527 TCTGACCTGCTGTGATCATCTCCACAAAGATGAATCATCCCGGGCTACTGACAG 586
DB 143004 TCTGACCTGCTGTGATCATCTCCACAAAGATGAATCATCCCGGGCTACTGACAG 143063
OY 587 CTCTGTGAGTCAACAGCCCTTTCAGTGGAGAACAGCGCTGCTGCTACAGCCAC 646
DB 143064 CTCTGTGAGTCAACAGCCCTTTCAGTGGAGAACAGCGCTGCTGCTACAGCCAC 143123
OY 647 GTGCTGTCAAGAGAGTGGCTTCTTCAACCTGACGCTCCCTGCTGCTGAAGAC 706
DB 143124 GTGCTGTCAAGAGAGTGGCTTCTTCAACCTGACGCTCCCTGCTGCTGAAGAC 143183
OY 707 CCTCTCTCAACCTGCTATGCTGACCTGTCCGGGACCCCGGGCGCTGTCCGCTTC 766
DB 143184 CCTCTCTCAACCTGCTATGCTGACCTGTCCGGGACCCCGGGCGCTGTCCGCTTC 143243
OY 767 CGAGAACGCACAAAGGAGATCTCATGATTGACAGTGTGATGGGGAGCATGAG 826
DB 143244 CGAGAACGCACAAAGGAGATCTCATGATTGACAGTGTGATGGGGAGCATGAG 143303
OY 827 CAAAACCTCAGAGAGAGACCAACCTCATATGATGAGTACAGTCTGCCAAAGCCAG 886
DB 143304 CAAAACCTCAGAGAGAGACCAACCTCATATGATGAGTACAGTCTGCCAAAGCCAG 143363
OY 887 CTGGGATGTCTAAGACATCCAGTCTTCCCAAGGCCCTGAGAAAGCTTACCTGCTT 946
DB 143364 CTGGGATGTCTAAGACATCCAGTCTTCCCAAGGCCCTGAGAAAGCTTACCTGCTT 143423
OY 947 GTGCGTATGAGAGTGTGCTGAGCCCTGTGGCCAGACATCCCGAATGATGAAATTC 1006
DB 143424 GTGCGTATGAGAGTGTGCTGAGCCCTGTGGCCAGACATCCCGAATGATGAAATTC 143483
OY 1007 GTGGGATGTGGAATTTCTGCCCATTTTCAGACCTGGGTCATTAATCAACCCGAGCAAG 1066
DB 143484 GTGGGATGTGGAATTTCTGCCCATTTTCAGACCTGGGTCATTAATCAACCCGAGCAAG 143543
OY 1067 GGCATGGGTGACAGGCTTTCACACAAATGCGAGGATGCCCTTAATGCTCCAGGCT 1126
DB 143544 GGCATGGGTGACAGGCTTTCACACAAATGCGAGGATGCCCTTAATGCTCCAGGCT 143603
OY 1127 TGGGCTGTGCTTTCCTTATGAAAAAGTTTCTCGACTTCAGAAAGCCGTGGCGATGCC 1186
DB 143604 TGGGCTGTGCTTTCCTTATGAAAAAGTTTCTCGACTTCAGAAAGCCGTGGCGATGCC 143663
OY 1187 ATGAATTTGCTGGCTACCGCAGCTGATGTAACAAGAACAGAAACCTGTTGCTG 1246
DB 143664 ATGAATTTGCTGGCTACCGCAGCTGATGTAACAAGAACAGAAACCTGTTGCTG 143723
OY 1247 GATCTTGTCTACTGAGCTGCTCCCTGAGCAAAATCCCTAAGAGGTTGAAAGGCTT 1306
DB 143724 GATCTTGTCTACTGAGCTGCTCCCTGAGCAAAATCCCTAAGAGGTTGAAAGGCTT 143783
OY 1307 GTGGCAGCTGTGTGACAGCTGATGATCTTCTGAAATGCTTGAAGCTTGGCTACAT 1366
DB 143784 GTGGCAGCTGTGTGACAGCTGATGATCTTCTGAAATGCTTGAAGCTTGGCTACAT 143843
OY 1367 CTCTGAGCTTAACTAATGATGCTGAGGATATCACTAGTGTGATGCTGCTCAACCT 1426
DB 143844 CTCTGAGCTTAACTAATGATGCTGAGGATATCACTAGTGTGATGCTGCTCAACCT 143903

```

```

OY 1427 GCTCAAGCAGAGGACTTTTGTGCTCATGCTTGTGTCTAAGAAACAGACTGGGAACTT 1486
DB 143904 GCTCAAGCAGAGGACTTTTGTGCTCATGCTTGTGTCTAAGAAACAGACTGGGAACTT 143963
OY 1487 ATGTAGCAGCAGATCCCAACAGTGAACAGGATATGCTCTTCTTCTTCTTATGATCT 1546
DB 143964 ATGTAGCAGCAGATCCCAACAGTGAACAGGATATGCTCTTCTTCTTCTTATGATCT 144023
OY 1547 CTTGTCTGGGAGACTTCAGACATTTTGGGCTGGAGGCTTATTAAGCAGACAGTA 1606
DB 144024 CTTGTCTGGGAGACTTCAGACATTTTGGGCTGGAGGCTTATTAAGCAGACAGTA 144083
OY 1607 TCAGTGAATTAATCAATTAACCTCCCTGTGTCACATCTTCCCAATGGGAAATGATCTT 1666
DB 144084 TCAGTGAATTAATCAATTAACCTCCCTGTGTCACATCTTCCCAATGGGAAATGATCTT 144143
OY 1667 TCACCAAGAGCTCAACAGCATTTTCCACAGAGATGC 1703
DB 144144 TCACCAAGAGCTCAACAGCATTTTCCACAGAGATGC 144180

RESULT 5
AC138848/c
LOCUS
DEFINITION
AC138848 Homo sapiens chromosome 5 clone RP11-1301B21, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
AC138848
AC138848.1 GI:27805260
VERSION
HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE
DOE Joint Genome Institute.
DOE Joint Genome Institute.
TITLE
Sequencing of Human Chromosome 5
JOURNAL
Unpublished
AUTHORS
2 (bases 1 to 206943)
DOE Joint Genome Institute.
DIRECT SUBMISSION
Submitted (21-JAN-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94596, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 2719263
Center clone name: RCT1-11_1301B21
-----
Summary Statistics
Consensus quality: 202039 bases at least Q40
Consensus quality: 202804 bases at least Q30
Consensus quality: 203337 bases at least Q20
Estimated insert size: 175000; agarose-fp estimation
Estimated insert size: 206343; sum-of-ctdigs estimation
Quality coverage: 15.28 in Q20 bases; agarose-fp estimation
Quality coverage: 12.96 in Q20 bases; sum-of-ctdigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
1
1220 1219: contig of 1219 bp in length
1320 1319: gap of unknown length
2493 2493: contig of 1174 bp in length
2494 2593: gap of unknown length
3666 3666: contig of 1073 bp in length
3667 3766: gap of unknown length

```


VERSION AK074746.1 GI:22760388
 KEYWORDS oligo capping; fls (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 REFERENCE 1
 AUTHORS Sugiyama, T., Ota, T., Nishikawa, T., Hayashi, K., Otsuki, T.,
 Kawai, H., O., Saito, K., Yamamoto, J., Wakamatsu, A., Nakamura, Y.,
 Kojima, S., Nagahara, K., Masuno, Y., Ono, T., Okano, K., Yoshikawa, Y.,
 Aotsuka, S., Sasaki, N., Hattori, A., Okumura, K., Iwayanagi, T. and
 Ninomiya, K.
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 1979)
 AUTHORS Isogai, T. and Otsuki, T.
 JOURNAL Direct Submission
 Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
 (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan: cDNA full insert sequencing;
 Research Association for Biotechnology: cDNA library construction;
 Institute of Medical Science, University of Tokyo, Laboratory of
 Genome Structure, Human Genome Center: cDNA 5'- & 3'-end one pass
 sequencing and clone selection: Helix Research Institute (supported
 by Japan Key Technology Center etc.).
 FEATURES
 SOURCE Location/Qualifiers
 1..1979
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2RM4.001325"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone_id="NT2RM4"
 /note="cloning vector: PME18SFL3-mRNA from uninduced NT2
 neuronal precursor cells"
 69..1229
 /note="unamed protein product"
 /codon_start=1
 /protein_id="BAC11177.1"
 /db_xref="GI:22760389"
 /translation="MLPFRKPKLFLVDSOMAILALFHHYSHNLSMKKAOERH
 VIASSMRSGSSFFGLFGHDPVFLVEMFAMVHWMFQSTAMLMHAYRDLIRAF
 LCDMSYVDAYNEBEPKROSSLFQWENRALSAPACITIPDEITIPAKHRLDSQP
 FEYERKCRNYSIVLKEVFEFNQSLVPLKDSLHLVHLVDRRAVRSRTERK
 GDLMDIRIVYGOHEKLEKEDQVLYMOVICOSQLHYTQSLIPRALDERVLYRY
 EDLARAFAVQTSNRYEFGLEFLPHLOTWVHNITRGGMGDHAFHTVARALNVSQAM
 RMSLPYKVKYRLOKACGDAMNLGLGRHVRSEQRNLILLDLSTWTVPEQIH"
 CDS
 ORIGIN
 BASE COUNT 451 a 542 c 474 g 512 t
 Query Match 75.7%; Score 1547; DB 9; Length 1979;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 346 GCCCGCTGACAGCTGTGATGACCTTCAAGCAGACGCGCTGATGCTGCATGCGC 405
 DB 287 GCCCGCTGACAGCTGTGATGACCTTCAAGCAGACGCGCTGATGCTGCATGCGC 346
 QY 406 TGTGCGGATCTGATACGCGCGCTCTTGTGGACATGAGCTCTTTGATGCTACAT 465
 DB 347 TGTGCGGATCTGATACGCGCGCTCTTGTGGACATGAGCTCTTTGATGCTACAT 406
 QY 466 GGAACCTGTGTCGCCGAGACAGTCCACCTCTTCACTGGAGAACAGCGCGCGTGTG 525
 DB 407 GGAACCTGTGTCGCCGAGACAGTCCACCTCTTCACTGGAGAACAGCGCGCGTGTG 466
 QY 526 TTTGTCACCTGCGCTGATATCGTACACCAAGATGAATATCCCGGGGCTACCTGAG 585
 DB 467 TTTGTCACCTGCGCTGATATCGTACACCAAGATGAATATCCCGGGGCTACCTGAG 526
 QY 586 GCTCTGTCAGTCAAGACGCTTTGAGTGTGAGAGAGCGCTGCTCTACAGCA 645
 DB 527 GCTCTGTCAGTCAAGACGCTTTGAGTGTGAGAGAGCGCTGCTCTACAGCA 586
 QY 646 CGTGTGCTCAAGAGGTGCGCTTCTTCAACCTGCACTCCCTTACCCGCTGAAAGA 705
 DB 587 CGTGTGCTCAAGAGGTGCGCTTCTTCAACCTGCACTCCCTTACCCGCTGAAAGA 646
 QY 706 CCCCTGCTCAACCTGATATCGTACACCTGCTGCGGAGACCGCGGGCGTTCGCTTC 765
 DB 647 CCCCTGCTCAACCTGATATCGTACACCTGCTGCGGAGACCGCGGGCGTTCGCTTC 706
 QY 766 CCGAGAACGCAAGAGGAGATCTCATGATTTGACAGTTCGATTTGATGGGCGACATGA 825
 DB 707 CCGAGAACGCAAGAGGAGATCTCATGATTTGACAGTTCGATTTGATGGGCGACATGA 766
 QY 826 GCAAAACTCAAGAGAGAGACCAACCTTACTATGATGAGTCACTTGTCCAAAGCA 885
 DB 767 GCAAAACTCAAGAGAGAGACCAACCTTACTATGATGAGTCACTTGTCCAAAGCA 826
 QY 886 GCTGAGATCTACAAAGACATCCAGTCTTCCCAAGGCGCTGAGAGAACGCTACCTGCT 945
 DB 827 GCTGAGATCTACAAAGACATCCAGTCTTCCCAAGGCGCTGAGAGAACGCTACCTGCT 886
 QY 946 TGTGCGCTGATGAGAGACCTGCTGAGCGCCCTGTGGCCAGACTTCCGAATGTGATTT 1005
 DB 887 TGTGCGCTGATGAGAGACCTGCTGAGCGCCCTGTGGCCAGACTTCCGAATGTGATTT 946
 QY 1006 CGTGGGATTTGGAATTTTGGCCCATCTTTCAGACCTGGGTGATTAATCAACCCGACGAA 1065
 DB 947 CGTGGGATTTGGAATTTTGGCCCATCTTTCAGACCTGGGTGATTAATCAACCCGACGAA 1006
 QY 1066 GGGCATGGGTCACACGCTTTCACACAAATGCCAGGAGTCCCTTAATGTCCTCCAGGC 1125
 DB 1007 GGGCATGGGTCACACGCTTTCACACAAATGCCAGGAGTCCCTTAATGTCCTCCAGGC 1066
 QY 1126 TTTGCGCTGTCTTTGGCCCATATGAAGATTTCTGACTTCAGAAACCTGTGGCGATTC 1185
 DB 1067 TTTGCGCTGTCTTTGGCCCATATGAAGATTTCTGACTTCAGAAACCTGTGGCGATTC 1126
 QY 1186 CATGAATTTGTGGGCTACACGCGCTTCCAGATCTGGAACAGAAAGAAACCTGTGCT 1245
 DB 1127 CATGAATTTGTGGGCTACACGCGCTTCCAGATCTGGAACAGAAAGAAACCTGTGCT 1186
 QY 1246 GGATCTTCTGTCACCTGAGATGTCCTGAGACAAATCCACTAAGAGGTTGAGAAGCTT 1305
 DB 1187 GGATCTTCTGTCACCTGAGATGTCCTGAGACAAATCCACTAAGAGGTTGAGAAGCTT 1246
 QY 1306 TGTGCGACCTGTGTGACGCTCAGTCACTTCTGTGAATGCTTCTGAGCCTTGCATACA 1365
 DB 1247 TGTGCGACCTGTGTGACGCTCAGTCACTTCTGTGAATGCTTCTGAGCCTTGCATACA 1306
 QY 1366 TCTCTGAGCCTTAACCTACATGTCGTGGGATATCACTAGATGATGATTTGTGTCAACG 1425
 DB 1307 TCTCTGAGCCTTAACCTACATGTCGTGGGATATCACTAGATGATGATTTGTGTCAACG 1366

QY 1426 TGTCAAGCAGAAAGACTTTTGTCTCAGTCTGTCTTAGAAAACAGACTGGGGAACCT 1485
DB 1367 TGTCAAGCAGAAAGACTTTTGTCTCAGTCTGTCTTAGAAAACAGACTGGGGAACCT 1426
QY 1486 TATGTAGCAGCAGATCCACCAGTGAACAGAGGTATGTCTTCTTCTTTCTTATCT 1545
DB 1427 TATGTAGCAGCAGATCCACCAGTGAACAGAGGTATGTCTTCTTCTTTCTTATCT 1486
QY 1546 TCTGTCTGTGGCAGACTTCAGAGACTTTGTGGCCTGAGAGCCCTATTAGCAGACAGT 1605
DB 1487 TCTGTCTGTGGCAGACTTCAGAGACTTTGTGGCCTGAGAGCCCTATTAGCAGACAGT 1546
QY 1606 ATCAGTGAATGTATCATTAACCTCCCTGTCTCAGACTTTGCCAATGGGGAATGATCT 1665
DB 1547 ATCAGTGAATGTATCATTAACCTCCCTGTCTCAGACTTTGCCAATGGGGAATGATCT 1606
QY 1666 TTCACCAAGAGCTCACAGACTTTTCCACAGAGATGC 1703
DB 1607 TTCACCAAGAGCTCACAGACTTTTCCACAGAGATGC 1644

RESULT 8
AR203335
LOCUS AR203335 2032 bp DNA linear PAT 20-JUN-2002
DEFINITION Sequence 2 from patent US 6365365.
ACCESSION AR203335
VERSION AR203335.1 GI:21499698
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistup,A., Rosen,S.D., Tangemann,K. and Hemmerich,S.
TITLES Method of determining whether an agent modulates glycosyl
sulfotransferase-3
JOURNAL Patent: US 6365365-A 2 02-APR-2002;
FEATURES Location/Qualifiers
source 1..2032
ORIGIN /organism="unknown"
BASE COUNT 468 a 569 c 490 g 505 t

Query Match 75.7%; Score 1547; DB 6; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Y 106 CAAGGCTTTCACATTCAGACAAATGCTACTGCTTAATAAATGAAGCTCTGCTTTCT 165
DB 160 CAAGGCTTTCACATTCAGACAAATGCTACTGCTTAATAAATGAAGCTCTGCTTTCT 219
QY 166 GGTTCCTCCAGATGGCATCTTGGCTTATTTTCCACATGTACAGCCACATCAGCTC 225
DB 220 GGTTCCTCCAGATGGCATCTTGGCTTATTTTCCACATGTACAGCCACATCAGCTC 279
QY 226 CCTGTCTATGAAGGACAGCCGAGGAGCATGACAGTGTCTGTCTTCTGGGGCTC 285
DB 280 CCTGTCTATGAAGGACAGCCGAGGAGCATGACAGTGTCTGTCTTCTGGGGCTC 339
QY 286 TGGCTCTTCTTTGTGGGAGAGCTTTTGGGAGCAGCCAGATGTTTCTTACCTGATGA 345
DB 340 TGGCTCTTCTTTGTGGGAGAGCTTTTGGGAGCAGCCAGATGTTTCTTACCTGATGA 399
QY 346 GCCCGCTGGCAGCTGTGATGACTTCAAGCAGAGACCCGCTGTGATCTGCACATGCG 405
DB 400 GCCCGCTGGCAGCTGTGATGACTTCAAGCAGAGACCCGCTGTGATCTGCACATGCG 459
QY 406 TGGGAGGATGTGATCGGGCGTCTTGTGTGCGCATGAGAGCTTGTGATGCGCTACAT 465
DB 460 TGTGCGGATCTGTATAGGGCGCTCTTGTGTGCGCATGAGAGCTTGTGATGCGCTACAT 519
QY 466 GGAAGCTGTGCTCCGAGAGACAGTCAAGCTTTTCAAGTGGGAGAACAGCCGGGCTGTG 525
DB 520 GGAAGCTGTGCTCCGAGAGACAGTCAAGCTTTTCAAGTGGGAGAACAGCCGGGCTGTG 579

QY 526 TTCTGACCTGCTGTGACATCATCCACAAAGATGAATCATCCCGGGGCTCACTGAG 585
DB 580 TTCTGACCTGCTGTGACATCATCCACAAAGATGAATCATCCCGGGGCTCACTGAG 639
QY 586 GCTCCGTGACAGTCAACACACCCCTTTAGAGTGTGTGAGAAAGGCTGCGCTCTACAGCA 645
DB 640 GCTCCGTGACAGTCAACACACCCCTTTAGAGTGTGTGAGAAAGGCTGCGCTCTACAGCA 699
QY 646 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGCAGTCCCTCTACCCCTGCTGAAGA 705
DB 700 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGCAGTCCCTCTACCCCTGCTGAAGA 759
QY 706 CCCCTCCCTCAACCTGCATATCGTGCACCTGTGCGGAGCCCGGGGCTGTTCGTTT 765
DB 760 CCCCTCCCTCAACCTGCATATCGTGCACCTGTGCGGAGCCCGGGGCTGTTCGTTT 819
QY 766 CCGAGAGAGCAGAAAGGAGATCTCATGTATGACATGCCATTTGTATGGGAGAGATGA 825
DB 820 CCGAGAGAGCAGAAAGGAGATCTCATGTATGACATGCCATTTGTATGGGAGAGATGA 879
QY 826 GCAAAACCTCAAGAAAGAGAGACCAACCTACTATGTATGACAGTATCTGCCAAAGCA 885
DB 880 GCAAAACCTCAAGAAAGAGAGACCAACCTACTATGTATGACAGTATCTGCCAAAGCA 939
QY 886 GCTGAGATCTACAAAGACATCCAGTCTTGGCCCAAGGCCCTGCAGAAAGCTTACCTGCT 945
DB 940 GCTGAGATCTACAAAGACATCCAGTCTTGGCCCAAGGCCCTGCAGAAAGCTTACCTGCT 999
QY 946 TGTGCGCTATGAGAGACCTGTGCTGAGCCCTGTGGGCGCAGACTTCCCGAATATGAAAT 1005
DB 1000 TGTGCGCTATGAGAGACCTGTGCTGAGCCCTGTGGGCGCAGACTTCCCGAATATGAAAT 1059
QY 1006 CGTGGATTTGAATTTCTCCCATCTTGTAGAGCTGTGGTCAATATCATCAGCCGAGCA 1065
DB 1060 CGTGGATTTGAATTTCTCCCATCTTGTAGAGCTGTGGTCAATATCATCAGCCGAGCA 1119
QY 1066 GGGCATGGTGTACACACGCTTTTCCACAAATGCGAGATGCCCTTAATGTCTCCAGGC 1125
DB 1120 GGGCATGGTGTACACACGCTTTTCCACAAATGCGAGATGCCCTTAATGTCTCCAGGC 1179
QY 1126 TTGGGCGTGTCTTCCCTATGAAAAGTTCTGACCTTCAGAAAAGCTGTGGGAGTC 1185
DB 1180 TTGGGCGTGTCTTCCCTATGAAAAGTTCTGACCTTCAGAAAAGCTGTGGGAGTC 1239
QY 1186 CATGAATTTGTGGGTACCGCAGCTGATCTGAAGAAGAACAGAGAACTGTGTCT 1245
DB 1240 CATGAATTTGTGGGTACCGCAGCTGATCTGAAGAAGAACAGAGAACTGTGTCT 1299
QY 1246 GGAATCTTGTCTACTGAGACTGTCCCTAGCAAAATCCACTAAGAGGTTGAGAGGCTT 1305
DB 1300 GGAATCTTGTCTACTGAGACTGTCCCTAGCAAAATCCACTAAGAGGTTGAGAGGCTT 1359
QY 1306 TGTGCGCACCTGTGTCAGCTCAGTCACTTTCTGATGAGTCTTGAGCCTTGCTTACA 1365
DB 1360 TGTGCGCACCTGTGTCAGCTCAGTCACTTTCTGATGAGTCTTGAGCCTTGCTTACA 1419
QY 1366 TCTGTGAGCTTAATCATATGCTGTGGGTATCACACTAGTGTGATGTGTCCACAG 1425
DB 1420 TCTGTGAGCTTAATCATATGCTGTGGGTATCACACTAGTGTGATGTGTGTCCACAG 1479
QY 1426 TGTCTAAGCAGAAAGACTTTTGTGTCTCATGCTTTGTCTTAGAAAACAGACTGGGAACCT 1485
DB 1480 TGTCTAAGCAGAAAGACTTTTGTGTCTCATGCTTTGTGTCTTAGAAAACAGACTGGGAACCT 1539
QY 1486 TATGTAGCAGCAGATCCACCAGTGAACAGAGGTATGTCTTCTTCTTTCTTATCT 1545
DB 1540 TATGTAGCAGCAGATCCACCAGTGAACAGAGGTATGTCTTCTTCTTTCTTATCT 1599
QY 1546 TCTGTCTGTGGCAGACTTCAGAGACTTTGTGGCCTGAGAGCCCTATTAGCAGACAGAT 1605
DB 1600 TCTGTCTGTGGCAGACTTCAGAGACTTTGTGGCCTGAGAGCCCTATTAGCAGACAGAT 1659

Db 1540 TATGAGCAGCAGCATCCACAGAGTATGCTCTTCTTTCTTTCTTGTATCT 1599

QY 1546 TCTGTCTGGGAGAGCTTTCAGAGACTTTTGCCCTGGAGCCATTATAGACAGACAGT 1605

Db 1600 TCTGTCTGGGAGAGCTTTCAGAGACTTTTGCCCTGGAGCCATTATAGACAGACAGT 1659

QY 1606 ATCAGTGAATGATCATTAACCTCCCTGTCACATCTTGCCCAATGGGGAATGATCT 1665

Db 1660 ATCAGTGAATGATCATTAACCTCCCTGTCACATCTTGCCCAATGGGGAATGATCT 1719

QY 1666 TTCACCAAGAGCTCCACAGCATTTTCCACAGAGATGC 1703

Db 1720 TTCACCAAGAGCTCCACAGCATTTTCCACAGAGATGC 1757

RESULT 10
AF131235 2032 bp mRNA linear PRI 10-AUG-1999
DEFINITION Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA,
complete cds.
ACCESSION AF131235
VERSION AF131235.1 GI:4927113
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE 1 (bases 1 to 2032)
Bistrup,A., Bhakta,S., Lee,J.-K., Belov,Y.Y., Gunn,M.D., Zuo,F.R.,
Huang,C.C., Kannagi,R., Rosen,S.D. and Hemmerich,S.
Sulfotransferases of two specificities function in the
reconstitution of high endothelial cell ligands for L-selectin
J. Cell Biol. 145 (4), 899-910 (1999)
JOURNAL 99264336
MEDLINE 10330415
PUBMED 2 (bases 1 to 2032)
Bistrup,A., Tangemann,K., Bhakta,S., Lee,J.-K., Belov,Y.Y.,
Gunn,M.D., Zuo,F.R., Huang,C.C., Kannagi,R., Rosen,S.D. and
Hemmerich,S.
Direct Submission
Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience,
3401 Hillview Avenue, Palo Alto, CA 94304, USA
FEATURES
source location/Qualifiers
1. 2032
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_type="high endothelial"
/tissue_type="tonsil"
182..1342
/function="biosynthesis of endothelial ligands for
L-selectin"
/note="HEC-G1cNAc-6-ST"
/codon_start=1
/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="AAD33015.1"
/db_xref="GI:4927114"
/translation="MLPRKMKLLFLVSQALILFFPMYSHNISLSMKQPERH
LVLSWSRSGSFFVQGLFGQHPDVYLMPPAWHYMTFKQSTAMMIMHAVRDLIRAV
LCDMSVFAIMRGPBRROSLFQWENSRALCSAPACDIPODEIIPRAICRLISQOP
FEVEKACRSYSHVYLKEVRFNLSLYLLKDPSLNLIHVLNDPRAVFRSREBK
GDLMDSRIVMGHOKLKKEDOPYVMVOICOSOLEIKTIOSLPKALOERYLLYRY
EDLARPAVAGTSRMKTFYVGLLEFLPHQITVHNITRKKGSDIAFHTNADALNVSQAW
RMSLEYEKSRQKACGDAMNLIGTRHVSQEDORNLIDLILSTVTPQIH"

BASE COUNT 468 a 569 c 490 g 505 t

ORIGIN

Query Match 75.7%; Score 1547; DB 9; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAGGCTTCCACTTCAGACATGCTACTGCTAAATAAAGAGCTCTGCTGTTTCT 165

Db 160 CAGGCTTCCACTTCAGACATGCTACTGCTAAATAAATAAAGAGCTCTGCTGTTTCT 219

QY 166 GGTTCACAGATGGCCATCTTGGCTATCTTCCACATGTACAGCACAACATAGCTC 225

Db 220 GGTTCACAGATGGCCATCTTGGCTATCTTCCACATGTACAGCACAACATAGCTC 279

QY 226 CCTGTCTATGAAGGACACAGCCGAGAGGCATGACGTGCTGTCTTCTCTGGCGCTC 285

Db 280 CCTGTCTATGAAGGACACAGCCGAGAGGCATGACGTGCTGTCTTCTCTGGCGCTC 339

QY 286 TGGCTTCTTTTGTGTGGGACACTTTTGGGACAGACCCAGATGTTTCTTACTGTATGA 345

Db 340 TGGCTTCTTTTGTGTGGGACACTTTTGGGACAGACCCAGATGTTTCTTACTGTATGA 399

QY 346 GCCCGCTGGACAGTGTGATGACCTTCAAGCAGACACCGCGTGGATGCTGCATATGC 405

Db 400 GCCCGCTGGACAGTGTGATGACCTTCAAGCAGACACCGCGTGGATGCTGCATATGC 459

QY 406 TGTGCGGATCTGATACGGGCGCTTCTTGTGCGACATGACCGTCTTGTATGCTTACAT 465

Db 460 TGTGCGGATCTGATACGGGCGCTTCTTGTGCGACATGACCGTCTTGTATGCTTACAT 519

QY 466 GGAACCTGTCGCCCGAGACAGTCCAGCGCTTTCAGTGGAGAACAGCCGGGCGCTGTG 525

Db 520 GGAACCTGTCGCCCGAGACAGTCCAGCGCTTTCAGTGGAGAACAGCCGGGCGCTGTG 579

QY 526 TTCTGCACCTGCTGTGATCATCTCCACAGATGAAATCATCCCGGGCTCACTGCAG 585

Db 580 TTCTGCACCTGCTGTGATCATCTCCACAGATGAAATCATCCCGGGCTCACTGCAG 639

QY 586 GCTCTGTGCTCAACACACCTTTGAGGTGTGGAAGAGCGCTGCGCTCTTACAGCA 645

Db 640 GCTCTGTGCTCAACACACCTTTGAGGTGTGGAAGAGCGCTGCGCTCTTACAGCA 699

QY 646 CGTGTGCTCAAGAGAGTGCCTTCTTCAACCTGCAGTCCCTCTACCGGCTGTAAGA 705

Db 700 CGTGTGCTCAAGAGAGTGCCTTCTTCAACCTGCAGTCCCTCTACCGGCTGTAAGA 759

QY 706 CCCCTCTCAACCTGCATATGCTGTGACCTGTGCGGAGCCCGGGCGTTCCTGTT 765

Db 760 CCCCTCTCAACCTGCATATGCTGTGACCTGTGCGGAGCCCGGGCGTTCCTGTT 819

QY 766 CCGAGAGCCACAAAGGAGATCTCATGTTGACAGTGCATTTGATGGGCGACATGA 825

Db 820 CCGAGAGCCACAAAGGAGATCTCATGTTGACAGTGCATTTGATGGGCGACATGA 879

QY 826 GCAAAAACACAAAGGAGAGACCAACCTACTATGATGACAGTCTATGCAAGCA 885

Db 880 GCAAAAACACAAAGGAGAGACCAACCTACTATGATGACAGTCTATGCAAGCA 939

QY 886 GCTGAGATCTACAGACCATCATGCTTGTGCCCAAGGCGCTGACAGAAAGCTACCTGCT 945

Db 940 GCTGAGATCTACAGACCATCATGCTTGTGCCCAAGGCGCTGACAGAAAGCTACCTGCT 999

QY 946 TGTGCGCTATGAGAGCCTGCTGAGCGCCCTGTGGCGCCAGACTTCCGAATGATGAAT 1005

Db 1000 TGTGCGCTATGAGAGCCTGCTGAGCGCCCTGTGGCGCCAGACTTCCGAATGATGAAT 1059

QY 1006 CGTGGATGATGATCTTGTGCCCATCTTCAGACCTGCGGTGATCAATCATCACCGAGCA 1065

Db 1060 CGTGGATGATGATCTTGTGCCCATCTTCAGACCTGCGGTGATCAATCATCACCGAGCA 1119

QY 1066 GGGCATGGGTGACACAGCTTTCACACAAATGCCAGAGATGCCCTTAATGTCTCCAGGC 1125

Db 1120 GGGCATGGGTGACACAGCTTTCACACAAATGCCAGAGATGCCCTTAATGTCTCCAGGC 1179

QY 1126 TTTGGCGCTGTGCTTGGCTTATGAAGAGTCTTGAGTTCAGAAAGCTGTGGGATGC 1185

Db 1180 TTTGGCGCTGTGCTTGGCTTATGAAGAGTCTTGAGTTCAGAAAGCTGTGGGATGC 1239

QY 1186 CATGAATTTGCTGGGCTACCGGACGTCGATCTGAACAAGAGAGAAACCTTGTCT 1245

Db 1240 CATGAATTTGCTGGGCTACCGGACGTCGATCTGAACAAGAGAGAAACCTTGTCT 1299

OY	1246	GGATCTCTGTCTACACGGAGCTGTGCCGAGCAAAATCCACTAAGAGGTTAGAAAGCTT	1305
Dp	1300	GGATCTTCTGTCTACCTGGACTGTCCCTGACCAATCCACTAAGAGGTTAGAAAGCTT	1359
OY	1306	TGCTGCCACCTGGTGTAGCCCTCAGTCAGTCTTCTGAAATGCTTCTGAGCCTTGCTTACA	1365
Dp	1360	TGCTGCCACCTGGTGTAGCCCTCAGTCAGTCTTCTGAAATGCTTCTGAGCCTTGCTTACA	1419
OY	1366	TCCTGAGCCTTAATCTACATGCTGTGGGTATCACACTGAGTGTGAGTTGTGTCCACAG	1425
Dp	1420	TCCTGAGCCTTAATCTACATGCTGTGGGTATCACACTGAGTGTGAGTTGTGTCCACAG	1479
OY	1426	TGCTCAAGCAGAAAGAGCTTTTGTGTCCATGCTTGTCTAGAAAAACAGACTGGGGAACCT	1485
Dp	1480	TGCTCAAGCAGAAAGAGCTTTTGTGTCCATGCTTGTCTAGAAAAACAGACTGGGGAACCT	1539
OY	1486	TATGTGAGCAGACATCCACACAGTGAACAGGGTATTTGCTCTTCTTCTTCTGTGATCT	1545
Dp	1540	TATGTGAGCAGACATCCACACAGTGAACAGGGTATTTGCTCTTCTTCTTCTGTGATCT	1599
OY	1546	TCTGTGTGGGCGACACTTCAGAGACTTTGTGGCCTGGAGGGCTTTTAAGCAGCGACAGT	1605
Dp	1600	TCTGTGTGGGCGACACTTCAGAGACTTTGTGGCCTGGAGGGCTTTTAAGCAGCGACAGT	1659
OY	1606	ATCAGTGGAAATGATTCATAAACCTGCCCTGTCCACATCTTCCCAATGGGGAAATGATGCT	1665
Dp	1660	ATCAGTGGAAATGATTCATAAACCTGCCCTGTCCACATCTTCCCAATGGGGAAATGATGCT	1719
OY	1666	TTTCACCAAAAGAGCTCACCGAGATTTTCCACAGAGATGC	1703
Dp	1720	TTTCACCAAAAGAGCTCACCGAGATTTTCCACAGAGATGC	1757

```

RESULT 11
AF149783
LOCUS AF149783 1333 bp mRNA linear PRI 02-JUL-2001
DEFINITION Homo sapiens L-selectin ligand sulfotransferase mRNA, complete cds.
ACCESSION AF149783
VERSION AF149783.1 GI:13897503
KEYWORDS
SOURCE
ORGANISM Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
TITLE
Yeh,J.C., Hiraoka,N., Petyniak,B., Nakayama,T., Ellies,L.G.,
Rabuke,D., Hindsгал,O., March,J.D., Lowe,J.B. and Fukuda,M.,
Novel sulfated lymphocyte homing receptors and their control by a
Core1 extension beta 1,3-N-acetylglucosaminyltransferase
Cell 105 (??), 957-969 (2001)
JOURNAL
MEDLINE 21332592
PUBMED 11439191
AUTHORS
2 (bases 1 to 1333)
Hiraoka,N. and Fukuda,M.
Direct Submission
Submitted (10-MAY-1999) Glycobiology, The Burnham Institute, 10901
North Torrey Pines Road, La Jolla, CA 92037, USA
FEATURES
SOURCE
location/qualifiers
1..1333
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
93..1253
/note="LSSP"
/codon_start=1
/product="L-selectin ligand sulfotransferase"
/protein_id="AAK48417.1"
/db_xref="GI:13897504"
/translation="MLPKKKMLLPVSONAIALPEHHYSHNISLSMKAQPERMH
VIVLSMRGSSSPFGOLFQHPDVFYALMHPAMHMYMFQSTAMNLMAYRDLRAVF
LCDSVDAWNEPGRQSSLCFQENSRALCSAPACITITODELTIPAHKRLICSDP
FEVEAKCRSTSHVLYEVEFNFQSTLYPLKDPISLNLIHVLIVDRPAAFRSRERK

```

Query Match	Best Local Similarity	Matches 1262: Conservative	Score 1262: 61.8%	Pred. No. 0: 0	Mismatches 0: 0	Indels 0: 0	Gaps 0: 0	Length 1333:
QY 107	AAAGTCTTCCACCTTCAGACAAATGCTACTGCGCTTAAAAAATGAAGCTCTGCTTCTG	166						
DB 72	AAGGTCTTCCACTTCAGACAAATGCTACTGCGCTTAAAAAATGAAGCTCTGCTTCTG	131						
OY 167	GTTTCCAGATGAGCCATCTTGCTCTATTCTTCCATATGACAGCCAAACATCAGCTCC	226						
DB 132	GTTTCCAGATGAGCCATCTTGCTCTATTCTTCCATATGACAGCCAAACATCAGCTCC	191						
OY 227	CTGTCTATGAAGGCACAGCCCGAGCGCATGACGTGCTGGTCTGTCTCTCTGCGCTCT	286						
DB 192	CTGTCTATGAAGGCACAGCCCGAGCGCATGACGTGCTGGTCTGTCTCTGCGCTCT	251						
OY 287	GGCTCTCTTTTGTGGGGCAGCTTTTGGGAGAGACCCAGATGTTTTCTTACCTATGAG	346						
DB 252	GGCTCTCTTTTGTGGGGCAGCTTTTGGGAGAGACCCAGATGTTTTCTTACCTATGAG	311						
OY 347	CCCGCTTGGCAGCTGTGATGACCTTCCAAAGCAGACACCGCTGATGCTGCATGAGCT	406						
DB 312	CCCGCTTGGCAGCTGTGATGACCTTCCAAAGCAGACACCGCTGATGCTGCATGAGCT	371						
OY 407	GTGGGGATCTGATACGGGGCGCTCTTCTTGTGCAATGAGCGTCTTGTATGCTTACATG	466						
DB 372	GTGGGGATCTGATACGGGGCGCTCTTCTTGTGCAATGAGCGTCTTGTATGCTTACATG	431						
OY 467	GAACTGGTCCCGGGAGACGATCCAGCGCTTTTAAAGGGGAGAACACCGGGGCGCTGCT	526						
DB 432	GAACTGGTCCCGGGAGACGATCCAGCGCTTTTAAAGGGGAGAACACCGGGGCGCTGCT	491						
OY 527	TCTGCACTGCTGTGACATCATCCCAAGATGAATATCAATCCCGGGGCTCACTGAGG	586						
DB 492	TCTGCACTGCTGTGACATCATCCCAAGATGAATATCAATCCCGGGGCTCACTGAGG	551						
OY 587	CTCCTGTGACATCAACAGCCCTTGTAGGTGTGAGAAAGGCTCGCTCCTTACAGCCAC	646						
DB 612	CTCCTGTGACATCAACAGCCCTTGTAGGTGTGAGAAAGGCTCGCTCCTTACAGCCAC	611						
OY 647	GTGGTGTCTAAGAGAGGTGCTGCTTCTTCAACCTGAGATCCCTTACCCGCTGTAAGAC	706						
DB 671	GTGGTGTCTAAGAGAGGTGCTGCTTCTTCAACCTGAGATCCCTTACCCGCTGTAAGAC	671						
OY 707	CCCTCCCTCAACCTGCATATGTGACACCTGTGCTCGGAGACCCCGGGGCGTTCCTGTC	766						
DB 731	CCCTCCCTCAACCTGCATATGTGACACCTGTGCTCGGAGACCCCGGGGCGTTCCTGTC	731						
OY 767	CGAGACGCAAAAGGAGATCTCATGATTTGACAGTGCATTTGTATGGGGCAGCATGAG	826						
DB 791	CGAGACGCAAAAGGAGATCTCATGATTTGACAGTGCATTTGTATGGGGCAGCATGAG	791						
OY 827	CAAAAACCTCAAGAGAGAGACCAACCCCTACTATGTATGCAAGTCACTGTGCAAAAGCCAG	886						
DB 851	CAAAAACCTCAAGAGAGAGACCAACCCCTACTATGTATGCAAGTCACTGTGCAAAAGCCAG	851						
OY 887	CTGAGATCTCTCAAGACCATTCAGTCTGCTGCTCGGACCCCTGACGAAAGCTTACTGCT	946						
DB 911	CTGAGATCTCTCAAGACCATTCAGTCTGCTGCTCGGACCCCTGACGAAAGCTTACTGCT	911						
OY 947	GTCGCGATATGAGACCTGGCTCGAGCCCTGTGTGGCCAGACTTCCGAAATGTAATTC	1006						
DB 971	GTCGCGATATGAGACCTGGCTCGAGCCCTGTGTGGCCAGACTTCCGAAATGTAATTC	971						
OY 1007	GTGGGATTTGGAATCTTGTCCCACTCTTCAAGACCTGGGTGCATATCACCAGGAGCAAG	1066						
DB 1031	GTGGGATTTGGAATCTTGTCCCACTCTTCAAGACCTGGGTGCATATCACCAGGAGCAAG	1031						

QY 1067 GGCATGGGTGACACGCTTTCCACAAATGCCAGGATGCCCTTAATGTCGCCAGGCT 1126
 DB 1032 GGCATGGGTGACACGCTTTCCACAAATGCCAGGATGCCCTTAATGTCGCCAGGCT 1091
 QY 1127 TGGCCCTGGTCTTTGCCCTATGAAAGGTTTCTGCACCTTCAGAAAGCCGTGGCAGTGC 1186
 DB 1092 TGGCCCTGGTCTTTGCCCTATGAAAGGTTTCTGCACCTTCAGAAAGCCGTGGCAGTGC 1151
 QY 1187 ATGAATTTGCTGGGTACCGCAGCTCAGATCTGAACAAGACAGAAAGCTGTGTCTG 1246
 DB 1152 ATGAATTTGCTGGGTACCGCAGCTCAGATCTGAACAAGACAGAAAGCTGTGTCTG 1211
 QY 1247 GATCTTCTGTCTACCTGACCTGCTCCCTGAGCAATTCACAAAGGTTGAGAAGGCTTT 1306
 DB 1212 GATCTTCTGTCTACCTGACCTGCTCCCTGAGCAATTCACAAAGGTTGAGAAGGCTTT 1271
 QY 1307 GGTGCACTGTGTGTCAGCTCAGTACCTTCTGTAATGCTTTCAGAGCTTGCCTACAT 1366
 DB 1272 GGTGCACTGTGTGTCAGCTCAGTACCTTCTGTAATGCTTTCAGAGCTTGCCTACAT 1331
 QY 1367 CT 1368
 DB 1332 CT 1333
 RESULT 12
 AR310485 2065 bp DNA linear PAT 12-JUN-2003
 LOCUS AR310485
 DEFINITION Sequence 26 from patent US 6558935.
 ACCESSION AR310485
 KEYWORDS AR310485.1 GI:31703448
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2065)
 AUTHORS Tang, Y. T., Corley, N. C., Guegler, K. J., Baughn, M. R., Lal, P., Yue, H.,
 Hillman, J. U., and Azimzal, Y.
 TITLE Human transferase-proteins
 JOURNAL Patent: US 6558935-A 26 06-MAY-2003;
 FEATURES
 source 1..2065
 ORIGIN /organism="Unknown"
 BASE COUNT 466 a 575 c 491 g 533 t
 Query Match 56.4%; Score 1153; DB 6; Length 2065;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

DB 452 TGTGGGGATGTGATAGGGCCGCTCTTGTGCGACATGAGCGTCTTGTATGCTACAT 511
 QY 466 GGAACCTGGTCCCGGAGACATCCAGCCTTTTCAGTGGAGAACAGCCGGCCCTGTG 525
 DB 512 GGAACCTGGTCCCGGAGACATCCAGCCTTTTCAGTGGAGAACAGCCGGCCCTGTG 571
 QY 526 TTCTGCACCTGGCTGTGACATCATCCACAGATGAATCATCCCGGGGCTCACTGCAG 585
 DB 572 TTCTGCACCTGGCTGTGACATCATCCACAGATGAATCATCCCGGGGCTCACTGCAG 630
 QY 586 GCTCTGTGACATCAAGAGCCCTTTG - AGSTGTGGAGAGGCGCCGCTCTACAGCC 644
 DB 631 GCTCTGTGACATCAAGAGCCCTTTGAGATTTTGGAGAAAGCTGCGCTCTACAGCC 690
 QY 645 AGTGTGTCTCAAGAGAGTGTGCTTCTCAACCTGAGTCCCTACCCGCTGTGAAAG 704
 DB 691 AGTGTGTCTCAAGAGAGTGTGCTTCTCAACCTGAGTCCCTACCCGCTGTGAAAG 750
 QY 705 ACCCTGCTCAACCTGCATATGCTGACCTGCTGCGGAGCCCGGGGCTGTCCGTT 764
 DB 751 ACCCTGCTCAACCTGCATATGCTGACCTGCTGCGGAGCCCGGGGCTGTCCGTT 810
 QY 765 CCCGAGAACGCAAGAGGAGATCATGATGACAGTGCATTTGTGATGGGCGAGCATG 824
 DB 811 CCCGAGAACGCAAGAGGAGATCATGATGACAGTGCATTTGTGATGGGCGAGCATG 870
 QY 825 AGCAAAATCTCAAGAGAGAGCAAACTTATGTATGTAGAGTCACTGCCAAAGCC 884
 DB 871 AGCAAAATCTCAAGAGAGAGCAAACTTATGTATGTAGAGTCACTGCCAAAGCC 930
 QY 885 AGCTGAGATCTCAAGAGAGAGCAAACTTATGTATGTAGAGTCACTGCCAAAGCC 944
 DB 931 AGCTGAGATCTCAAGAGAGAGCAAACTTATGTATGTAGAGTCACTGCCAAAGCC 990
 QY 945 TTGTGCGATATGAGAGAGCTGCTGAGCCCTGTGCGGAGAGCTTCCGAAATGTAAAT 1004
 DB 991 TTGTGCGATATGAGAGAGCTGCTGAGCCCTGTGCGGAGAGCTTCCGAAATGTAAAT 1050
 QY 1005 TGTGGGATGTGAATTTTGTCCCATCTTTCAGACCTGGGTGATTAACATCCCGAGCA 1064
 DB 1051 TGTGGGATGTGAATTTTGTCCCATCTTTCAGACCTGGGTGATTAACATCCCGAGCA 1110
 QY 1065 AGGCAATGGGTACACAGCTTTCACACAAAGCCAGAGATGCCCTTATGTCTCCAGG 1124
 DB 1111 AGGCAATGGGTACACAGCTTTCACACAAAGCCAGAGATGCCCTTATGTCTCCAGG 1170
 QY 1125 CTGGGCGTGTGCTTGTCCCTATGAAAGGTTTCTGCACTTCAAGAAAGCCGTGGCAGT 1184
 DB 1171 CTGGGCGTGTGCTTGTCCCTATGAAAGGTTTCTGCACTTCAAGAAAGCCGTGGCAGT 1230
 QY 1185 CCATGAATTTGCTGGGCTACCGCAGCTCAGATCTGAACAGAGAGAAAGCTGTGTC 1244
 DB 1231 CCATGAATTTGCTGGGCTACCGCAGCTCAGATCTGAACAGAGAGAAAGCTGTGTC 1290
 QY 1245 TGGATCTTGTCTACCTGTGATGCTGCTGAGCAAAATCAGTAAAGAGGTATGAAGGT 1304
 DB 1291 TGGATCTTGTCTACCTGTGATGCTGCTGAGCAAAATCAGTAAAGAGGTATGAAGGT 1350
 QY 1305 TTGTCGACCTGTGTGTCAGCTCAGTCACTTCTGTAATCTTCTGAGCCCTGCTAC 1364
 DB 1351 TTGTCGACCTGTGTGTCAGCTCAGTCACTTCTGTAATCTTCTGAGCCCTGCTAC 1410
 QY 1365 ATCTGTAGACCTTAACATCATGCTGTGGGTATCAACATGATGTGATGTGTCCACAC 1424
 DB 1411 ATCTGTAGACCTTAACATCATGCTGTGGGTATCAACATGATGTGATGTGTCCACAC 1470
 QY 1425 GTGCTCAAGCAAGAGACTTTTGTGTCCATGCTTGTGTACAAAACAGACTGGGAAAC 1484
 DB 1471 GTGCTCAAGCAAGAGACTTTTGTGTCCATGCTTGTGTACAAAACAGACTGGGAAAC 1530
 QY 1485 TTATGTAGACACACATCCACAGTGAAGAGAGGATTTGCTTCTTCTTGTGATC 1544
 DB 1531 TTATGTAGACACACATCCACAGTGAAGAGAGGATTTGCTTCTTCTTGTGATC 1590

OY 1545 TTCCGTCTGGGAGACTTTCAGAGACTTTGGCTGGAGGCTATTAAAGCAGACAG 1604
 DB 1591 TTCCGTCTGGGAGACTTTCAGAGACTTTGGCTGGAGGCTATTAAAGCAGACAG 1650
 OY 1605 TATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTCCCAATGGGAATGATC 1664
 DB 1651 TATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTCCCAATGGGAATGATC 1710
 OY 1665 TTTCCACCAAGAGCTCAGCAGCATTTTCCACAGAGATGC 1703
 DB 1711 TTTCCACCAAGAGCTCAGCAGCATTTTCCACAGAGATGC 1749
 RESULT 13
 BD124832
 LOCUS BD124832 877 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD124832
 VERSION BD124832.1 GI:23219777
 KEYWORDS JP 2002017375-A/263
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
 Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
 Koga,H.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 263 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/263
 PD 22-JAN-2002
 PE 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10,
 PC C12P21/02,C12O1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT source 1..877
 /organism="Homo sapiens (human)"
 /location="1..877
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 183 a 266 c 221 g 203 t 4 others
 ORIGIN
 Query Match 34.0%; Score 695; DB 6; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 OY 106 CAAAGTCTTCACCTTCAGACAAATGCTACTGCTTAAATAAATGAAGCTCTGCTTTCT 165
 DB 47 CAAAGTCTTCACCTTCAGACAAATGCTACTGCTTAAATAAATGAAGCTCTGCTTTCT 106
 OY 166 GGTTCACAGATGCCCATCTGCTATTCTTCCACATGTACAGCCACAATCAGCTC 225
 DB 107 GGTTCACAGATGCCCATCTGCTATTCTTCCACATGTACAGCCACAATCAGCTC 166
 OY 226 CCGTGTATGAAGCAGACCCGAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 285
 DB 167 CCGTGTATGAAGCAGACCCGAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 226

OY 286 TGGCTCTTTTGTGGGAGACTTTTGGGAGACACCAGATGTTTCTACATGATGA 345
 DB 227 TGGCTCTTTTGTGGGAGACTTTTGGGAGACACCAGATGTTTCTACATGATGA 286
 OY 346 GCCCGCCGAGCAGTGTGATGATCTTCCACAGACAGCAGCCGCTGATGCTGACATGCG 405
 DB 287 GCCCGCCGAGCAGTGTGATGATCTTCCACAGACAGCAGCCGCTGATGCTGACATGCG 346
 OY 406 TGTGGGATCTGATACGGCCGCTCTTTGTGGACATGAGCGCTTTGATGCTTACAT 465
 DB 347 TGTGGGATCTGATACGGCCGCTCTTTGTGGACATGAGCGCTTTGATGCTTACAT 406
 OY 466 GGAACCTGTCGCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACACCGGGCCCTGTG 525
 DB 407 GGAACCTGTCGCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACACCGGGCCCTGTG 466
 OY 526 TTTGACACCTGCTGTGACATCATCCCAAGATGAATATCCCGCGGCTCATGTCAG 585
 DB 467 TTTGACACCTGCTGTGACATCATCCCAAGATGAATATCCCGCGGCTCATGTCAG 526
 OY 586 GCTGCTGTGACATCAACAGCCCTTGAAGTGTGAGAGAGCGCTGCTGCTACAGCA 645
 DB 527 GCTGCTGTGACATCAACAGCCCTTGAAGTGTGAGAGAGCGCTGCTGCTACAGCA 586
 OY 646 CGTGAGTCAAGAGAGTGGCTCTTCAACCTGACAGTCCCTACCGCTGCTGAAGA 705
 DB 587 CGTGAGTCAAGAGAGTGGCTCTTCAACCTGACAGTCCCTACCGCTGCTGAAGA 646
 OY 706 CCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGCGCTGCTGCTG 765
 DB 647 CCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGCGCTGCTGCTG 706
 OY 766 CCGAGAGCAGCAAGAGAGATCTCATATTGACA 800
 DB 707 CCGAGAGCAGCAAGAGAGATCTCATATTGACA 741
 RESULT 14
 BD126950
 LOCUS BD126950 877 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD126950
 VERSION BD126950.1 GI:23221895
 KEYWORDS JP 2002017375-A/2381.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
 Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
 Koga,H.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 2381 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/2381
 PD 22-JAN-2002
 PE 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10,
 PC C12P21/02,C12O1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT source 1..877
 /organism="Homo sapiens (human)"
 /location="1..877
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 183 a 266 c 221 g 203 t 4 others
 ORIGIN
 Query Match 34.0%; Score 695; DB 6; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 FEATURES
 TITLE
 JOURNAL
 COMMENT
 ORIGIN
 Query Match 34.0%; Score 695; DB 6; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 FEATURES
 TITLE
 JOURNAL
 COMMENT
 ORIGIN

source 1.877
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 183 a 266 c 221 g 203 t 4 others
 ORIGIN

Query Match 34.0%; Score 695; DB 6; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 CAGGCTTCACCTTACGACAAATGCTACTGCTATAAAAAAATCAACCTCTCTGCTTCT 165
 |||||||
 Db 47 CAGGCTTCACCTTACGACAAATGCTACTGCTATAAAAAAATCAACCTCTCTGCTTCT 106
 |||||||
 QY 166 GGTTCACAGATGGCCATCTTGCTATCTTCCACATGTACAGCACAACATCAGCTC 225
 |||||||
 Db 107 GGTTCACAGATGGCCATCTTGCTATCTTCCACATGTACAGCACAACATCAGCTC 166
 |||||||
 QY 226 CCTGTATGAAAGGACACAGCCGAGCGCATGACGTGTCTGTCTCTGCGGCTC 285
 |||||||
 Db 167 CCTGTATGAAAGGACACAGCCGAGCGCATGACGTGTCTGTCTCTGCGGCTC 226
 |||||||
 QY 286 TGGCTCTCTTGTGGGCGAGCTTTTGGGCGACACCGAGATGTTTCTACTGTATGA 345
 |||||||
 Db 227 TGGCTCTCTTGTGGGCGAGCTTTTGGGCGACACCGAGATGTTTCTACTGTATGA 286
 |||||||
 QY 346 GCCCGCTTGGCAGATGTGATGACCTTCAAGCAGACACCGCTGATGCTGACATGGC 405
 |||||||
 Db 287 GCCCGCTTGGCAGATGTGATGACCTTCAAGCAGACACCGCTGATGCTGACATGGC 346
 |||||||
 QY 406 TGTGCGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTGTATGCTTACAT 465
 |||||||
 Db 347 TGTGCGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTGTATGCTTACAT 406
 |||||||
 QY 466 GGAACCTGTGCCCGGAGACAGTCGACCTTTCAGTGGGAGAAAGCGGCGCTGTG 525
 |||||||
 Db 407 GGAACCTGTGCCCGGAGAGTCAGCTCTTTCAGTGGGAGAAAGCGGCGCTGTG 466
 |||||||
 QY 526 TTCTGACCTGTGCTGATCATATCCACAGATGAATATATCCCGGGGCTCACTGACG 585
 |||||||
 Db 467 TTCTGACCTGTGCTGATCATATCCACAGATGAATATATCCCGGGGCTCACTGACG 526
 |||||||
 QY 586 GCTCCTGTGACATCAACACCCCTTTGAGGTGTGGAGAAAGCGCTGCTCTACAGCA 645
 |||||||
 Db 527 GCTCCTGTGACATCAACACCCCTTTGAGGTGTGGAGAAAGCGCTGCTCTACAGCA 586
 |||||||
 QY 646 CGTGTGCTCAAGGAGGTGCGCTTCTTCAACCGTCAGTCCCTACCGCTGCTGAAGA 705
 |||||||
 Db 587 CGTGTGCTCAAGGAGGTGCGCTTCTTCAACCGTCAGTCCCTACCGCTGCTGAAGA 646
 |||||||
 QY 706 CCCCTCCCTCAACCTGATATGCTGACCTGTGCGGACCCCGGGCGGTGTCGCTTC 765
 |||||||
 Db 647 CCCCTCCCTCAACCTGATATGCTGACCTGTGCGGACCCCGGGCGGTGTCGCTTC 706
 |||||||
 QY 766 CCGAGAACGCAAAAGGAGATCTCATGATTGACA 800
 |||||||
 Db 707 CCGAGAACGCAAAAGGAGATCTCATGATTGACA 741
 |||||||

RESULT 15
 AX381256 517 bp DNA linear PAT 18-MAR-2002
 LOCUS AX381256
 DEFINITION Sequence 194 from Patent WO212280.
 ACCESSION AX381256
 VERSION AX381256.1 GI:19576075
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 Pyle, R.A., Xu, J. and Secrist, H.

TITLE Compositions and methods for the therapy and diagnosis of colon
 cancer
 JOURNAL Patent: WO 0212280-A 194 14-FEB-2002;
 CORIXA CORPORATION (US)
 FEATURES location/Qualifiers
 source 1.517
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

BASE COUNT 121 a 130 c 124 g 142 t
 ORIGIN

Query Match 19.7%; Score 403; DB 6; Length 517;
 Best Local Similarity 100.0%; Pred. No. 3.5e-228;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCTTGGCCACCGCTGCTGACGCTCAGTCACTTCTCTGAATGCTTGAAGCTTGC 1360
 |||||||
 Db 13 GGCTTGGCCACCGCTGCTGACGCTCAGTCACTTCTCTGAATGCTTGAAGCTTGC 72
 |||||||
 QY 1361 CTACATCTGAGCCTTAATCAATGCTGTGGGTATCACACTGAGTGTGATGCTC 1420
 |||||||
 Db 73 CTACATCTGAGCCTTAATCAATGCTGTGGGTATCACACTGAGTGTGATGCTC 132
 |||||||
 QY 1421 ACACGTGCTCAAGCAGAGACTTTGTGTCCATGCTGTCTAGAAAAACAGACTGGG 1480
 |||||||
 Db 133 ACACGTGCTCAAGCAGAGACTTTGTGTCCATGCTGTCTAGAAAAACAGACTGGG 192
 |||||||
 QY 1481 AACCTTATGAGCGACATCCACAGTGAAGAGGATGCTCTCTCTTCTT 1540
 |||||||
 Db 193 AACCTTATGAGCGACATCCACAGTGAAGAGGATGCTCTCTCTTCTT 252
 |||||||
 QY 1541 GATCTTCTGTGCGCAGACTTTCAGAGACTTGTGGCTGAGAGGCTTATTAAGCAGAC 1600
 |||||||
 Db 253 GATCTTCTGTGCGCAGACTTTCAGAGACTTGTGGCTGAGAGGCTTATTAAGCAGAC 312
 |||||||
 QY 1601 ACAGTATCAAGTGAATGATTCATTAACCTCCCTGTCCACATCTTGGCCATGGGAATG 1660
 |||||||
 Db 313 ACAGTATCAAGTGAATGATTCATTAACCTCCCTGTCCACATCTTGGCCATGGGAATG 372
 |||||||
 QY 1661 GATCTTTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
 |||||||
 Db 373 GATCTTTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 415
 |||||||

Search completed: August 15, 2003, 03:08:35
 Job time : 7285 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:37:47 ; Search time 544 Seconds

(without alignments)
10137.779 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043
Sequence: 1 gaattccattgttggta.....tggatccctaaaaa 2043

Scoring table: OLIGO_NTC
Gapco 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2674488

Minimum DB seq length: 50

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*

1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1547	75.7	1979	22	AAK94229 Human full-length
2	1547	75.7	2032	20	AA220792 Human glycosyl su
3	1262	61.8	1333	24	AA516947 Human I-selectin s
4	1153	56.4	2065	21	AA294211 Human transferrase
5	695	34.0	877	22	AAK91803 Human CDNA 5'-end
6	695	34.0	877	22	AAK93921 Human CDNA clone I
7	403	19.7	505	24	ABW89280 Human colon cancer
8	403	19.7	517	24	ABK54724 Human colon cancer

9	306	15.0	2988	21	AACT6156 Human OREF1711
10	114	5.6	389	24	ABK54794 Human colon cancer
11	105	5.1	548	22	AAK92588 Human CDNA 3'-end
12	59	2.9	1647	24	AA024670 Human drug metabol
13	59	2.9	1694	22	AA02700 Human glycosyl su
14	59	2.9	2044	22	AA02699 Human glycosyl su
15	59	2.9	2170	22	AA02699 Human glycosyl su
16	59	2.9	2544	24	ABN89506 Human cornel N-ac
17	59	2.9	48436	24	ABN89533 Human cornel N-ac
18	59	2.9	160552	22	AA02697 Human glycosyl su
19	27	1.3	1926	20	AA220793 Mouse glycosyl su
20	23	1.1	1937	24	AA516948 Murine intestinal-
21	23	1.1	1989	22	AA02696 Mouse glycosyl su
22	22	1.1	167	24	ABL86480 Human ovarian can
23	22	1.1	171	24	ABL86089 Human ovarian can
24	22	1.1	187	24	ABL86207 Human ovarian can
25	22	1.1	191	24	ABL86306 Human ovarian can
26	21	1.0	10004	22	ABAI4483 Human nervous syst
27	21	1.0	10246	22	ABAI4485 Human nervous syst
28	20	1.0	88	21	AA025980 Human secreted pro
29	20	1.0	354	22	AA05813 Novel human polynu
30	20	1.0	370	22	AA05341 Novel human polynu
31	20	1.0	381	22	AA05789 Novel human polynu
32	20	1.0	398	25	ABX45934 Bovine EST associa
33	20	1.0	3763	19	AAV58194 Human myosin I-cha
34	20	1.0	3763	23	AA569872 DNA encoding novel
35	20	1.0	5814	24	ABN87855 Human ovary specif
36	20	1.0	6147	25	ACC46453 Human dlthp protei
37	20	1.0	9516	22	AAK7355 Human immune/haema
38	20	1.0	9516	22	AAK81394 Human immune/haema
39	20	1.0	13192	22	AAK73556 Human immune/haema
40	20	1.0	13192	22	AAK81396 Human immune/haema
41	20	1.0	35959	22	AAK78275 Human immune/haema
42	20	1.0	180557	22	ABN85750 Human BAC clone RP
43	20	1.0	349980	22	AA06431 Pyrococcus abyssi
44	19	0.9	341	24	ABL85863 Human ovarian can
45	19	0.9	341	24	ABL85910 Human ovarian can

ALIGNMENTS

RESULT 1	AAK94229	standard; CDNA: 1979 BP.
ID	AAK94229	
AC	AAK94229	
DT	06-NOV-2001	(first entry)
XX		
DE	Human full-length CDNA, SEQ ID NO: 2816.	
XX		
KW	Human: full length CDNA: CDNA synthesis; oligo-capping; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	EP1130094-A2.	
XX		
PD	05-SEP-2001.	
XX		
PF	07-JUL-2000; 2000EP-0114089.	
XX		
PR	08-JUL-1999; 99JP-0194486.	
PR	11-JAN-2000; 2000JP-0118774.	
PR	02-MAY-2000; 2000JP-0183765.	
XX		
PA	(HELI-) HELIX RES INST.	
PI	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;	
PI	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;	
XX		
DR	WPI: 2001-524255/58.	
DR	P-PSDB: AAM93309.	

XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
XX
PS Claim 8: SEQ ID NO 2816; 1380bp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5' - and 3' - ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other;
XX
Query Match 75.7%; Score 1547; DB 22; Length 1979;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
XX
QY 106 CAGGCTCTTCAGCTTACGACATGCTACTGCTTAAATAATGAAAGCTGCTGTTCT 165
DB 47 CAGGCTCTTCAGCTTACGACATGCTACTGCTTAAATAATGAAAGCTGCTGTTCT 106
QY 166 GGTTCGCCAGATGGGCGATCTGGCTATCTTCCAGATGTCAGCCAAATCAGCTC 225
DB 107 GGTTCGCCAGATGGGCGATCTGGCTATCTTCCAGATGTCAGCCAAATCAGCTC 166
QY 226 CCGTCTTAAAGAGCAGACCCGAGGCGATGCACTGCTGCTTCTTCTGCGGCTC 285
DB 167 CCGTCTTAAAGAGCAGACCCGAGGCGATGCACTGCTGCTTCTTCTGCGGCTC 226
QY 286 TGGCTCTTCTTTTGGGGGAGCTTTTGGGAGCAGCCAGATGTTTCTTACCTGATGA 345
DB 227 TGGCTCTTCTTTTGGGGGAGCTTTTGGGAGCAGCCAGATGTTTCTTACCTGATGA 286
QY 346 GCCCGCTGCGACAGTGTGATGACCTTCAAGAGAGACACCGCTGGATCTGACATGAGC 405
DB 287 GCCCGCTGCGACAGTGTGATGACCTTCAAGAGAGACACCGCTGGATCTGACATGAGC 346
QY 406 TGTGCGGAGATGTGATACGGGCGGTCTTGTGCGACATGAGCGTCTTGAATGCTACAT 465
DB 347 TGTGCGGAGATGTGATACGGGCGGTCTTGTGCGACATGAGCGTCTTGAATGCTACAT 406
QY 466 GGAACCTGTCCCGGAGACATCCAGCTCTTTCAGTGGGAGAAAGCCGGCCCTGTG 525
DB 407 GGAACCTGTCCCGGAGACATCCAGCTCTTTCAGTGGGAGAAAGCCGGCCCTGTG 466
QY 526 TTCTGACACTGCTGTGACATCATCCACAAATGAAATCATCCCGGGGCTCAGTCAG 585
DB 467 TTCTGACACTGCTGTGACATCATCCACAAATGAAATCATCCCGGGGCTCAGTCAG 526
QY 586 GCTCTGTGACATCAACAGCCCTTGAAGGTGTGAAGAGCCCTGCGCTCTACAGCA 645
DB 527 GCTCTGTGACATCAACAGCCCTTGAAGGTGTGAAGAGCCCTGCGCTCTACAGCA 586
QY 646 GGTGAGTCAAGAGGAGGTGCGCTTCAACCTGCAAGTCCCTACCCGCTGCTGAAGA 705
DB 587 GGTGAGTCAAGAGGAGGTGCGCTTCAACCTGCAAGTCCCTACCCGCTGCTGAAGA 646
QY 706 CCCCTCCCTCAACATGATATGTCGACCTGTCGCGAGCCCGGGGCGGTGTCGTTTC 765
DB 647 CCCCTCCCTCAACATGATATGTCGACCTGTCGCGAGCCCGGGGCGGTGTCGTTTC 706
QY 766 CCGAAGACGACAAAGGAGATCTCATGATGACAGTGCATGTTGATGGGAGCATGA 825
DB 707 CCGAAGACGACAAAGGAGATCTCATGATGACAGTGCATGTTGATGGGAGCATGA 766

QY 826 GCATAAACTCAAGAGAGGAGACCACTACTATGTGATGACAGTCTGCCAAGCCA 885
DB 767 GCATAAACTCAAGAGAGGAGACCACTACTATGTGATGACAGTCTGCCAAGCCA 826
QY 886 GCTGGAGATCTACAAACATCATGCTTGGCCAAAGGCCCTGACGAACCTACTGCT 945
DB 827 GCTGGAGATCTACAAACATCATGCTTGGCCAAAGGCCCTGACGAACCTACTGCT 886
QY 946 TGGGCTTAAAGAGAGCTGCTGAGCCCTGAGCCAGACTTCCGAATGATGAAT 1005
DB 887 TGGGCTTAAAGAGAGCTGCTGAGCCCTGAGCCAGACTTCCGAATGATGAAT 946
QY 1006 CGTGGATGGAATTTCTCCCATCTTCAAGCTGCTGATGATACATCAGCCAGGCA 1065
DB 947 CGTGGATGGAATTTCTCCCATCTTCAAGCTGCTGATGATACATCAGCCAGGCA 1006
QY 1066 GGGCATGGGTGACACAGCTTTCCACAAATGCGAGGATGCCCTTAATGTCGCCAGC 1125
DB 1007 GGGCATGGGTGACACAGCTTTCCACAAATGCGAGGATGCCCTTAATGTCGCCAGC 1066
QY 1126 TGGGCGCTGCTTGGCCCTATGAAAGGTTTCTGACTTCAGAAAGCCGTGGGATGC 1185
DB 1067 TGGGCGCTGCTTGGCCCTATGAAAGGTTTCTGACTTCAGAAAGCCGTGGGATGC 1126
QY 1186 CATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAACAAGAACAGAAACCTGTGCT 1245
DB 1127 CATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAACAAGAACAGAAACCTGTGCT 1186
QY 1246 GATCTTCTGTCTACTGAGCTGCTCCCTGAGCAAACTCAAGAGGTTGAGAAAGCTT 1305
DB 1187 GATCTTCTGTCTACTGAGCTGCTCCCTGAGCAAACTCAAGAGGTTGAGAAAGCTT 1246
QY 1306 TGTGCGACCTGGGTGACAGCTCAGCTCAGTCTTCTGATGATGTTCTGAGCCTTGCCTACA 1365
DB 1247 TGTGCGACCTGGGTGACAGCTCAGCTCAGTCTTCTGATGATGTTCTGAGCCTTGCCTACA 1306
QY 1366 TCTCTGAGCTTAACTAACTATGCTGTGGGTATCACAGTGTGATGATGCTCCACAGC 1425
DB 1307 TCTCTGAGCTTAACTAACTATGCTGTGGGTATCACAGTGTGATGATGCTCCACAGC 1366
QY 1426 TGTCTAAGCAGAGAGCTTGTGATGATGCTGTGATGATGATGATGATGATGATGAT 1485
DB 1367 TGTCTAAGCAGAGAGCTTGTGATGATGCTGTGATGATGATGATGATGATGATGAT 1426
QY 1486 TATGTAGAGCAGACATCCACAGTGAAGAGGATGATGATGATGATGATGATGATGAT 1545
DB 1427 TATGTAGAGCAGACATCCACAGTGAAGAGGATGATGATGATGATGATGATGATGAT 1486
QY 1546 TCTCTGTCTGGGAGACTTGTGAGAGACTTGTGGCTGTGAGAGGCTATTAAACAGACACAT 1605
DB 1487 TCTCTGTCTGGGAGACTTGTGAGAGACTTGTGGCTGTGAGAGGCTATTAAACAGACACAT 1546
QY 1606 ATCAGTGAATGATGATCAATCAATCCCTGTCACATCTTGGCCAAATGGGAGATGAT 1665
DB 1547 ATCAGTGAATGATGATCAATCAATCCCTGTCACATCTTGGCCAAATGGGAGATGAT 1606
QY 1666 TTACCAAAAGAGCTACAGCAATTTTCCACAGAGATGC 1703
DB 1607 TTACCAAAAGAGCTACAGCAATTTTCCACAGAGATGC 1644

RESULT 2
AAZ20792
ID AAZ20792 standard; DNA; 2032 BP.
XX
XX AAZ20792;
AC
AC 08-DEC-1999 (first entry)
DT
DT Human glycosyl sulfotransferase-3 coding sequence.
XX
XX Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
XX selectin binding interaction; inflammation; lymphocyte homing; human;
KW

KW secondary lymph organ; ss.
XX Homo sapiens.
XX W09949018-A1.
XX 30-SEP-1999.
XX 26-FEB-1999; 99WO-US04316.
XX 20-MAR-1998; 98US-0045284.
XX 12-NOV-1998; 98US-0190911.
XX (REGC) UNIV CALIFORNIA.
XX (SYNT) SYNTAX USA INC.
XX Blstrup A, Rosen SD, Tangemann K, Hemmerlich S;
XX WPI; 1999-580442/49.
XX P-PSDB; AAY39918.
XX Human and murine glycosyl sulfotransferase 3 and related
XX polynucleotides
XX
XX Claim 4; Fig 1; 59pp; English.
XX
XX This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of
XX the invention. The nucleic acid sequences, probes and primers derived
XX from these, proteins and antibodies are useful in detecting homologues.
XX The sequences, antibodies and methods are useful in the diagnosis and
XX treatment of diseases associated with selectin binding interactions,
XX including conditions associated with or resulting from the homing of
XX leukocytes to sites of inflammation and the normal homing of lymphocytes
XX to secondary lymph organs.
XX
XX Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other;
SQ
Query Match 75.7%; Score 1547; DB 20; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 106 CAAGGCTCTCCATGACACACATGCTACTGCTCTTAAATGAAGCTCTGCTGTTTCT 165
DB 160 CAAGGCTCTCCATGACACACATGCTACTGCTCTTAAATGAAGCTCTGCTGTTTCT 219
QY 166 GGTTCACAGATGGGCAATCTTGCTATTTCTTCCACATGACACCAACATCAGCTC 225
DB 220 GGTTCACAGATGGGCAATCTTGCTATTTCTTCCACATGACACCAACATCAGCTC 279
QY 226 CCTGCTATGAAAGCAGACAGCCGAGCGATGACGTGCTGCTTCTTCTGCGCTC 285
DB 280 CCTGCTATGAAAGCAGACAGCCGAGCGATGACGTGCTGCTTCTTCTGCGCTC 339
QY 286 TGCGCTCTTCTTGTGGGAGCTTTTGGGAGACACCCAGATGTTTCTTCACTGATGA 345
DB 340 TGCGCTCTTCTTGTGGGAGCTTTTGGGAGACACCCAGATGTTTCTTCACTGATGA 399
QY 346 GCCCGCTGACGCTGAGATGACCTTCAACAGACAGCCGCTGAGTGTGACATGGC 405
DB 400 GCCCGCTGACGCTGAGATGACCTTCAACAGACAGCCGCTGAGTGTGACATGGC 459
QY 406 TGTGCGGAGATGATGAGGCGCTTCTTGTGAGATGAGCTTCTTGTGATGCTACAT 465
DB 460 TGTGCGGAGATGATGAGGCGCTTCTTGTGAGATGAGCTTCTTGTGATGCTACAT 519
QY 466 GAAAGCTGTGCTCCGGAGACAGTCCAGCTCTTCACTGAGAGAACAGCCGGCTGTG 525
DB 520 GAAAGCTGTGCTCCGGAGACAGTCCAGCTCTTCACTGAGAGAACAGCCGGCTGTG 579
QY 526 TTCGACACCTGCTGATGATCATTCCACAGATGAATCATCCCGGGGCTCACTGAG 585
DB 580 TTCGACACCTGCTGATGATCATTCCACAGATGAATCATCCCGGGGCTCACTGAG 639

QY 586 GCTCCTGTGACATCAACAGCCCTTGTAGGTGTGAGAGAGCTGCGCTCTCAACGCA 645
DB 640 GCTCCTGTGACATCAACAGCCCTTGTAGGTGTGAGAGAGCTGCGCTCTCAACGCA 699
QY 646 CGTGGTCTCAAGAGAGTGGCTTCTTCAACCTGACGTCCCTTACCCGCTGTAAGA 705
DB 700 CGTGGTCTCAAGAGAGTGGCTTCTTCAACCTGACGTCCCTTACCCGCTGTAAGA 759
QY 706 CCCCTCCCTCAACCTGCAATATGTCGACCTGCTGCGGAGACCCCGGGGCGTCCGCTC 765
DB 760 CCCCTCCCTCAACCTGCAATATGTCGACCTGCTGCGGAGACCCCGGGGCGTCCGCTC 819
QY 766 CCGAGAAAGCAGAAAGGAGATCTCATGATTTGACAGTGCATTTGTGATGGGAGCATGA 825
DB 820 CCGAGAAAGCAGAAAGGAGATCTCATGATTTGACAGTGCATTTGTGATGGGAGCATGA 879
QY 826 GCAAAAGCTCAAGAGAGAGACCAACCTTATGTATGACAGTGCATTTGCAAGGCA 885
DB 880 GCAAAAGCTCAAGAGAGAGACCAACCTTATGTATGACAGTGCATTTGCAAGGCA 939
QY 886 GCTGAGATCTCAAGAGACATCCAGTCTTGGCCCAAGGCGCTGAGAGAGCTACCTGCT 945
DB 940 GCTGAGATCTCAAGAGACATCCAGTCTTGGCCCAAGGCGCTGAGAGAGCTACCTGCT 999
QY 946 TGTGCGCTATGAGAGACCTGCTGAGCCCTGTGGCCAGACTTCCCGAATGATGAAT 1005
DB 1000 TGTGCGCTATGAGAGACCTGCTGAGCCCTGTGGCCAGACTTCCCGAATGATGAAT 1059
QY 1006 CGTGGATTTGAAATTTTGGCCCATCTTCAAGACCTGGTGCATTAACATACCCAGGCA 1065
DB 1060 CGTGGATTTGAAATTTTGGCCCATCTTCAAGACCTGGTGCATTAACATACCCAGGCA 1119
QY 1066 GGGCATGGGTACCAACGCTTTCCACACAAAGCCAGAGATCCCTTATGTCTCCAGG 1125
DB 1120 GGGCATGGGTACCAACGCTTTCCACACAAAGCCAGAGATCCCTTATGTCTCCAGG 1179
QY 1126 TTGGGCTGCTCTTGTGCTTGAAGAGTTTCTGATGAGAAAGCCTGTGGAGATG 1185
DB 1180 TTGGGCTGCTCTTGTGCTTGAAGAGTTTCTGATGAGAAAGCCTGTGGAGATG 1239
QY 1186 CATGAATTTGCTGGGCTACCGCCACGTGATCTGACACAGACAGAAACCTGTGCT 1245
DB 1240 CATGAATTTGCTGGGCTACCGCCACGTGATCTGACACAGACAGAAACCTGTGCT 1299
QY 1246 GGATCTTCTGCTACCTGAGACGCTCCGAGACAAATCAGTAAGGGGTGGAAGGCT 1305
DB 1300 GGATCTTCTGCTACCTGAGACGCTCCGAGACAAATCAGTAAGGGGTGGAAGGCT 1359
QY 1306 TGTGCGACCTGTGTGACAGCCCTAGTCACTTCTTCTGAAATGCTTGTGAGCCCTGCTACA 1365
DB 1360 TGTGCGACCTGTGTGACAGCCCTAGTCACTTCTTCTGAAATGCTTGTGAGCCCTGCTACA 1419
QY 1366 TGTGCGACCTTAACTACATGCTGTGGGTATACACTGAGTGTGATGTGTCACAG 1425
DB 1420 TGTGCGACCTTAACTACATGCTGTGGGTATACACTGAGTGTGATGTGTCACAG 1479
QY 1426 TGTGCGACAGAAAGACTTGTGTGACATGCTTGTGTGATGAAAAACAGAGGGGAACCT 1485
DB 1480 TGTGCGACAGAAAGACTTGTGTGACATGCTTGTGTGATGAAAAACAGAGGGGAACCT 1539
QY 1486 TATGTGAGCAGACATCCACAGAGTGAAGAGGATATGCTTCTTCTTGTGATCT 1545
DB 1540 TATGTGAGCAGACATCCACAGAGTGAAGAGGATATGCTTCTTCTTGTGATCT 1599
QY 1546 TCTGTGCTGGGAGACTTCAAGAGACTTGTGGCTGAGGCGCTATTAAGCAGACAGT 1605
DB 1600 TCTGTGCTGGGAGACTTCAAGAGACTTGTGGCTGAGGCGCTATTAAGCAGACAGT 1659
QY 1606 ATCAGTGAATTTGATTCATTAACCTCCCTGTGACATCTTCCCAATGGGGAATGATCT 1665
DB 1660 ATCAGTGAATTTGATTCATTAACCTCCCTGTGACATCTTCCCAATGGGGAATGATCT 1719
QY 1666 TTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1703

Db 1720 TTCACCAAGAGCTCACGACATTTTCACAGAGATGC 1757

RESULT 3

AA16947 standard; cDNA; 1333 BP.

AA16947;

12-MAR-2002 (first entry)

Human L-selectin sulfotransferase-2 (LST-2) cDNA.

Human: beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;
L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
delayed-type hypersensitivity reaction; hyperplastic thymus; antileuker;
antiinflammatory; antipsoriatic; antidiabetic; dermatological;
antiallergic.

Homo sapiens.

Location/Qualifiers

111..1253

/tag= a

/product= "Human LST-2"

WO200185177-A1.

15-NOV-2001.

10-MAY-2001; 2001WO-US15452.

11-MAY-2000; 2000US-0569320.

(BURN-) BURHAM INST.

Fukuda M, Yeh J, Hirakawa N;

WPI; 2002-075226/10.

P-PSDB; AAU11274.

New enzyme, useful for modifying acceptor molecule, comprises an isolated L-selectin sulfotransferase-2 that directs expression of L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal GlcNAc 6-sulfotransferase

Claim 19; Fig 4; 98pp; English.

The present invention provides a method of modifying an acceptor molecule by contacting the acceptor with an isolated beta1,3-N-acetylglucosaminyltransferase (beta1,3GNT) or an active fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The invention also provides a method of treating or preventing an L-selectin-mediated condition by reducing the expression or activity of a beta1,3GNT that directs expression of a MECA-79 antigen. This can be done by administering to the subject an oligosaccharide L-selectin antagonist that inhibits binding of L-selectin to a MECA-79 antigen, for example by administering antibody material that specifically binds beta1,3GNT, and/or a beta1,3GNT antisense nucleic acid molecule. L-selectin sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression. Alternatively, the expression or activity of LST-2 or its active fragment can be reduced in combination with reducing the expression or activity of beta1,3GNT. The method is useful for treating L-selectin mediated conditions such as Crohn's disease and ulcerative colitis, inflammatory disorders of the skin such as allergic contact dermatitis, psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type hypersensitivity reactions, diabetes and hyperplastic thymus. This sequence represents cDNA encoding human LST-2.

Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;

Query Match 61.8%; Score 1262; DB 24; Length 1333;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	107	AAGCTTCCACCTTCACGACATGCTACTGCTTAAATAAAGAGCTCTGCTGTTCTG	166
DB	72	AAGCTTCCACCTTCACGACATGCTACTGCTTAAATAAAGAGCTCTGCTGTTCTG	131
QY	167	GTTTCCAGATGGCGATCTGGCTATCTTCACATGATGACGCAACATCAGCTCC	226
DB	132	GTTTCCAGATGGCGATCTGGCTATCTTCACATGATGACGCAACATCAGCTCC	191
QY	227	CTGTCTATGAGGACGACAGCCGAGCCATGACAGCTGCTGTTCTGCTTCTGCGCTCT	286
DB	192	CTGTCTATGAGGACGACAGCCGAGCCATGACAGCTGCTGTTCTGCTTCTGCGCTCT	251
QY	287	GCTCTCTCTTTTGTGGGACGCTTTTGGGACACCCAGATGTTTCTAAGTATGAG	346
DB	252	GCTCTCTCTTTTGTGGGACGCTTTTGGGACACCCAGATGTTTCTAAGTATGAG	311
QY	347	CCGCTGACGACGTGTGATGACCTTCAAGCAGACGACGCGTGTGATGCTGACATGGCT	406
DB	312	CCGCTGACGACGTGTGATGACCTTCAAGCAGACGACGCGTGTGATGCTGACATGGCT	371
QY	407	GTCGGGATCTGATACGGGCGCTCTTCTGTGCGACATGACGCTTTGTATGCTTACATG	466
DB	372	GTCGGGATCTGATACGGGCGCTCTTCTGTGCGACATGACGCTTTGTATGCTTACATG	431
QY	467	GAACCTGGTCCCGGAGACAGTCACGCTCTTCAAGGAGAAAGCCGGGCGCTGTGT	526
DB	432	GAACCTGGTCCCGGAGACAGTCACGCTCTTCAAGGAGAAAGCCGGGCGCTGTGT	491
QY	527	TCTGACATGCTCTGTGATCATCCACAGATGAATATATCCCGGGCTCACTGACAG	586
DB	492	TCTGACATGCTCTGTGATCATCCACAGATGAATATATCCCGGGCTCACTGACAG	551
QY	587	CTCTGTGACATGAACAGCCCTTGTAGGTGTGAGAAAGCCCTGCTCTTACAGCCAC	646
DB	552	CTCTGTGACATGAACAGCCCTTGTAGGTGTGAGAAAGCCCTGCTCTTACAGCCAC	611
QY	647	GTCGTCTCAAGAGGAGGCGCTTCTCAACCTGACGCTGACCTGACCGGCTGAAAGAC	706
DB	612	GTCGTCTCAAGAGGAGGCGCTTCTCAACCTGACGCTGACCTGACCGGCTGAAAGAC	671
QY	707	CCCTCCCTCAACCTGATATGTCGACCTGTGTCGGGACCCCGGGCGCTGTCCTCC	766
DB	672	CCCTCCCTCAACCTGATATGTCGACCTGTGTCGGGACCCCGGGCGCTGTCCTCC	731
QY	767	CGAGAACGCAAAAGGAGATCTCATGATTTGACATGCGCATTTGTATGGGGACGATGAG	826
DB	732	CGAGAACGCAAAAGGAGATCTCATGATTTGACATGCGCATTTGTATGGGGACGATGAG	791
QY	827	CAAAAACGCAAAAGGAGGACCAACCTCATATGATGATGAGGTCATGCGCAAGACAG	886
DB	792	CAAAAACGCAAAAGGAGGACCAACCTCATATGATGATGAGGTCATGCGCAAGACAG	851
QY	887	CTGAGATCTACAAAGACATCCAGCTTGTGCCCAAGGCCCTGACGAGAACGCTACGCTT	946
DB	852	CTGAGATCTACAAAGACATCCAGCTTGTGCCCAAGGCCCTGACGAGAACGCTACGCTT	911
QY	947	GTCGCTATGAGAGCTGTGCTGAGGCGCTGTGCGCCAGACTTCCGGAATGTATGAAATTC	1006
DB	912	GTCGCTATGAGAGCTGTGCTGAGGCGCTGTGCGCCAGACTTCCGGAATGTATGAAATTC	971
QY	1007	GTCGGAATGGAATTTCTGGCCCATCTTCAGACCTGGGTGATTAATCAACCCGAGGCAAG	1066
DB	972	GTCGGAATGGAATTTCTGGCCCATCTTCAGACCTGGGTGATTAATCAACCCGAGGCAAG	1031
QY	1067	GGCATGGGTGACACGCTTTCCACACAATGCGAGGATGCTTATGTCTTCCAGGCT	1126
DB	1032	GGCATGGGTGACACGCTTTCCACACAATGCGAGGATGCTTATGTCTTCCAGGCT	1091

QY 1127 TGCGGCTGCTCTTGGCCCTATGAAAGTTTCTGACTTCGAAAGCCGTGGCGATGCC 1186
 DB 1092 TGGCGCTGCTCTTGGCCCTATGAAAGTTTCTGACTTCGAAAGCCGTGGCGATGCC 1151
 QY 1187 ATGAATTTGCTGGGCTACCCGACAGTCTGATCTGAACAGACAGAAACCTTGTCTG 1246
 DB 1152 ATGAATTTGCTGGGCTACCCGACAGTCTGATCTGAACAGACAGAAACCTTGTCTG 1211
 QY 1247 GATCTTGTCTGACTTCGAGCTGCTGAGCAATTCAGTAAAGGGGTGAGAGGCTTT 1306
 DB 1212 GATCTTGTCTGACTTCGAGCTGCTGAGCAATTCAGTAAAGGGGTGAGAGGCTTT 1271
 QY 1307 GCTGCCACCTGTGTCTGACCTCAGTCACTTCTCTGATGCTTGTGAGCCCTGCTACAT 1366
 DB 1272 GCTGCCACCTGTGTCTGACCTCAGTCACTTCTCTGATGCTTGTGAGCCCTGCTACAT 1331
 QY 1367 CT 1368
 DB 1332 CT 1333

RESULT 4
 AA294211
 ID AA294211 standard; cDNA; 2065 BP.
 AC AA294211;
 XX 19-JUN-2000 (first entry)
 DE Human transferase TRNSFS-11 cDNA clone 2617407CBI.
 XX
 XX Transferase; TRNSFS-11; human; antitumour; cell proliferation;
 KM inflammation; gastrointestinal disorder; developmental disorder;
 KM genetic disorder; neurological disorder; reproductive disorder;
 KM smooth muscle disorder; immunological disorder; gene therapy;
 KM diagnosis; N-acetylglucosamine 6-O-sulfotransferase; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 174..1334
 FT /*CDS- a
 XX
 PN WO200014251-A2.
 XX
 PD 16-MAR-2000.
 XX
 PF 09-SEP-1999; 99WO-US20989.
 XX
 PR 10-SEP-1998; 98US-0150657.
 PR 04-NOV-1998; 98US-0186779.
 PR 11-MAY-1999; 99US-0133642.
 XX
 PA (INCYT-) INCYTE PHARM INC.
 PI Tang YT, Corley NC, Guegler KJ, Baughn MR, Lai P, Yue H;
 PI Hillman JL, Azimzal Y;
 XX
 DR WPI: 2000-256996/22.
 DR P-PSDB: AAY79219.
 XX
 PT Human transferase proteins useful for preventing, diagnosing and
 PT treating cancers and developmental, gastrointestinal, genetic,
 PT immunological, neurological, reproductive and smooth muscle disorders -
 XX
 PS Claim 9; Page 104-105; 113pp; English.
 CC
 CC The present sequence is that of cDNA clone 2617407CBI encoding
 CC human transferase TRNSFS-11 (see AAY79219), 1 of 15 claimed human
 CC transferase proteins of the invention (see AAY79209-23). The clone
 CC was isolated from gall bladder cDNA library GBIANT01. TRNSFS-11
 CC is expressed in dermatologic and gastrointestinal tissues,
 CC especially those associated with inflammation and cell

CC proliferation. It shows homology to mouse N-acetylglucosamine
 CC 6-O-sulfotransferase. A polynucleotide comprising nucleotides
 CC 264-323 or 1272-1331 of the present sequence can be used as a DNA
 CC probe. The new human transferases and polynucleotides can be used
 CC in the diagnosis, prevention and treatment (including gene therapy
 CC and antisense therapy) of cancer, developmental disorders,
 CC gastrointestinal disorders, genetic disorders, immunological
 CC disorders, neurological disorders, reproductive disorders, and
 CC smooth muscle disorders.
 XX
 SO Sequence 2065 BP; 466 A; 575 C; 491 G; 533 T; 0 other.

Query Match 56.4%; Score 1153; DB 21; Length 2065;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 106 CAAAGCTTCCACTTTCAGACACAAATGCTACGCTTAAAAAATGAAAGCTCTGCTTTCT 165
 DB 152 CAAAGCTTCCACTTTCAGACACAAATGCTACGCTTAAAAAATGAAAGCTCTGCTTTCT 211
 QY 166 GGTTCACAGATGGCCATCTTGGCTATTTCTTCACATGTACAGCCACAATACAGCTC 225
 DB 212 GGTTCACAGATGGCCATCTTGGCTATTTCTTCACATGTACAGCCACAATACAGCTC 271
 QY 226 CCGTCTATGAAAGCACAGCCCGAGCGATGACGTGCTGCTGCTCTTCTGCGGCTC 285
 DB 272 CCGTCTATGAAAGCACAGCCCGAGCGATGACGTGCTGCTGCTCTTCTGCGGCTC 331
 QY 286 TGAGCTCTTCTTTTGGGGGAGCTTTTGGGACACCCAGATGTTTCTTACATGATGA 345
 DB 332 TGAGCTCTTCTTTTGGGGGAGCTTTTGGGACACCCAGATGTTTCTTACATGATGA 391
 QY 346 GCCCGCTGGACCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 405
 DB 392 GCCCGCTGGACCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATG 451
 QY 406 TGTGCGGAGATCTGATGAGGCGCGCTCTTGTGCGACATGAGCTTTGATGCTACAT 465
 DB 452 TGTGCGGAGATCTGATGAGGCGCGCTCTTGTGCGACATGAGCTTTGATGCTACAT 511
 QY 466 GGAACCTGTGCTCCCGGAGACAGTCCAGCTTTTCACTGAGGAGAAACCGGCGCTGTG 525
 DB 512 GGAACCTGTGCTCCCGGAGACAGTCCAGCTTTTCACTGAGGAGAAACCGGCGCTGTG 571
 QY 526 TTTGCACTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 585
 DB 572 TTTGCACTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 630
 QY 586 GCTCCTGTGAGTCAAGAGCCCTTTG-AGTGTGTGAGAAAGGCTGCGCTCTACAGCC 644
 DB 631 GCTCCTGTGAGTCAAGAGCCCTTTGAAAGTTGTGAGAAAGGCTGCGCTCTACAGCC 690
 QY 645 ACGTGTGCTCAAGAGAGTGGCTTTTCAACTGCACTGAGTCCCTTACCCGCTGTGAAG 704
 DB 691 ACGTGTGCTCAAGAGAGTGGCTTTTCAACTGCACTGAGTCCCTTACCCGCTGTGAAG 750
 QY 705 ACCCTCCCTCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 764
 DB 751 ACCCTCCCTCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 810
 QY 765 CCCGAGAAAGCAAGAGGAGATCTGATGATGATGATGATGATGATGATGATGATGATGATG 824
 DB 811 CCCGAGAAAGCAAGAGGAGATCTGATGATGATGATGATGATGATGATGATGATGATGATG 870
 QY 825 AGCAAAATCTCAAGAGAGAGCAACCTCTATGTGTGATGATGATGATGATGATGATGATG 884
 DB 871 AGCAAAATCTCAAGAGAGAGCAACCTCTATGTGTGATGATGATGATGATGATGATGATG 930
 QY 885 AGCTGAGATCTTACAAAGACCTCAGTCTTGGCCAAAGGCGCTCAGAAAGCTACCTGC 944
 DB 931 AGCTGAGATCTTACAAAGACCTCAGTCTTGGCCAAAGGCGCTCAGAAAGCTACCTGC 990
 QY 945 TTGTGCGCTATGAGAGCTGTGCTGAGACCCCTGTGTGCGCCAGACTTCCCGAATGTATGANT 1004

```

Db      991 TTTGGGCTATGAGGACCTGGCTCGAGCCCTGTGGCCAGACTTCCGGAATATGAAAT 1050
QY      1005 TCGTGGATTTGGAATCTTGTCCCATCTTCAGACCTGGGTGCAATACATCACCGAGCA 1064
Db      1051 TCGTGGATTTGGAATCTTGTCCCATCTTCAGACCTGGGTGCAATACATCACCGAGCA 1110
QY      1065 AGGGCATGGGTGACACCGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGG 1124
Db      1111 AGGGCATGGGTGACACCGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGG 1170
QY      1125 CTGGGCGTGTGCTTGTGGCCCTATGAAAGGTTCTGACCTTGAAAGGCTGTGGGATG 1184
Db      1171 CTGGGCGTGTGCTTGTGGCCCTATGAAAGGTTCTGACCTTGAAAGGCTGTGGGATG 1230
QY      1185 CCATGAATTTGTGGGCTACCGCCAGCTGAGATCTGACAGAAAGAGAAACCTGTGGC 1244
Db      1231 CCATGAATTTGTGGGCTACCGCCAGCTGAGATCTGACAGAAAGAGAAACCTGTGGC 1290
QY      1245 TGGATCTTGTCTTACTGAGACGTGCTCCAGCAATTCACATGAAGGGTTGGAAGGCT 1304
Db      1291 TGGATCTTGTCTTACTGAGACGTGCTCCAGCAATTCACATGAAGGGTTGGAAGGCT 1350
QY      1305 TTGCTGCACTGTGTGACAGCTCAGTCACTTCTCTGAATGCTTCTGAGCCCTTGCTAC 1364
Db      1351 TTGCTGCACTGTGTGACAGCTCAGTCACTTCTCTGAATGCTTCTGAGCCCTTGCTAC 1410
QY      1365 ATCTGAGCCCTTAATACATGCTGTGGGTATCAGACATGATGTGATGTGTCCACAC 1424
Db      1411 ATCTGAGCCCTTAATACATGCTGTGGGTATCAGACATGATGTGATGTGTCCACAC 1470
QY      1425 GTGCTCAGCAGAAAGGACTTTGTGTCATGCTGTGTGAGAAACAGACGTGGGAAAC 1484
Db      1471 GTGCTCAGCAGAAAGGACTTTGTGTCATGCTGTGTGAGAAACAGACGTGGGAAAC 1530
QY      1485 TATGTGACAGACATCCACCAAGTGAAGAAAGGATTTGCTCTTCTTCTTGATC 1544
Db      1531 TATGTGACAGACATCCACCAAGTGAAGAAAGGATTTGCTCTTCTTCTTGATC 1590
QY      1545 TTCTCTGTGGGAGACTTCAGAGACTTTGTGGCTGTGAGGCTTAAAGACAGACAG 1604
Db      1591 TTCTCTGTGGGAGACTTCAGAGACTTTGTGGCTGTGAGGCTTAAAGACAGACAG 1650
QY      1605 TATCAGTGAATTTGATCCATAAACCCTGTCATCATCTTGCCCAATGGGGAATGGATC 1664
Db      1651 TATCAGTGAATTTGATCCATAAACCCTGTCATCATCTTGCCCAATGGGGAATGGATC 1710
QY      1665 TTTCACCAAGAGCTCAGCAGATTTTCACAGAGATGC 1703
Db      1711 TTTCACCAAGAGCTCAGCAGATTTTCACAGAGATGC 1749

RESULT 5
AAK91803 standard; cDNA; 877 BP.
XX
AC AAK91803;
XX
06-NOV-2001 (first entry)
XX
DE Human cDNA 5'-end sequence, SEQ ID NO: 263.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS Homo sapiens.
XX
PN EP130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000BP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.

```

```

PR      11-JAN-2000; 2000BP-0118774.
PR      02-MAY-2000; 2000BP-0183765.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI; 2001-524255/58.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 2; SEQ ID NO 263; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 5'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;
XX
Query Match 34.0%; Score 695; DB 22; Length 877;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
QY      106 CAGGCTTCCACTTCAGACATGCTACTGCTTAAAAAATGAACTCTGCTTTCT 165
Db      47 CAAGGCTTCCACTTCAGACATGCTACTGCTTAAAAAATGAACTCTGCTTTCT 106
QY      166 GGTTCCTCCAGAGGCGCATCTTGCTATTTCTTCACATGATACGCCACATCAGCTC 225
Db      107 GGTTCCTCCAGAGGCGCATCTTGCTATTTCTTCACATGATACGCCACATCAGCTC 166
QY      226 CCGTCTATGAAGGCGACCGGAGCGCATGCGATGCGTCTCTCTCTGGGCGTC 285
Db      167 CCGTCTATGAAGGCGACCGGAGCGCATGCGATGCGTCTCTCTCTGGGCGTC 226
QY      286 TGGCTCTTCTTTGTGGGAGAGCTTTTGGGAGCAGCCAGATGTTTCTACTGATGGA 345
Db      227 TGGCTCTTCTTTGTGGGAGAGCTTTTGGGAGCAGCCAGATGTTTCTACTGATGGA 286
QY      346 GCCCGCTGGCAGCGTGTGATGACTTTCAGCAGACACCGCTGGATCTGCATATGC 405
Db      287 GCCCGCTGGCAGCGTGTGATGACTTTCAGCAGACACCGCTGGATCTGCATATGC 346
QY      406 TGTGCGGATCTGATACGGGCGCTCTTCTTGTGCGCATGACGCTTTTGATGCTTACT 465
Db      347 TGTGCGGATCTGATACGGGCGCTCTTCTTGTGCGCATGACGCTTTTGATGCTTACT 406
QY      466 GGAACCTGTCTCCCGGAGACAGTCAGCTCTTTTCTAGTGGAGAACAGCGGCGCTGTG 525
Db      407 GGAACCTGTCTCCCGGAGACAGTCAGCTCTTTTCTAGTGGAGAACAGCGGCGCTGTG 466
QY      526 TTCTGACCTGCTGTGATCATCCACAGATGAATCATCCCGGGGCTCACTGCAG 585
Db      467 TTCTGACCTGCTGTGATCATCCACAGATGAATCATCCCGGGGCTCACTGCAG 526
QY      586 GCTCTGTGCTCAACAGACCCCTTTGAGTGTGAGAAAGGCGCTGCGCTCTTAACGCA 645
Db      527 GCTCTGTGCTCAACAGACCCCTTTGAGTGTGAGAAAGGCGCTGCGCTCTTAACGCA 586
QY      646 CGTGGTCTCAAGGAGGTGCGCTTCTTCAACCTGCAAGTCCCTACCCGCTCTTAAGA 705
Db      587 CGTGGTCTCAAGGAGGTGCGCTTCTTCAACCTGCAAGTCCCTACCCGCTCTTAAGA 646

```


XX New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer
 XX
 PS Claim 1: SEQ ID NO 2595; 266bp + Sequence Listing: English.
 CC
 CC The invention relates to a human colon tumour expressed polynucleotide
 CC (I) encoding a polypeptide (II), ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i) under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (i) and (ii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pat_sequences.

XX Sequence 505 BP; 120 A; 125 C; 119 G; 141 T; 0 other;

XX Query Match 19.7%; Score 403; DB 24; Length 505;

XX Best Local Similarity 100.0%; Pred. No. 2.6e-189;

XX Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1301 GGCTTGGTGCACACCTGCTGTCAGCTCAGTCACTTCTCTGAAATGCTTCTGAGCCTTGC 1360

XX 1 GGCTTGGTGCACACCTGCTGTCAGCTCAGTCACTTCTCTGAAATGCTTCTGAGCCTTGC 60

XX 1361 CTACATCTCTGAGACCTTAACATACATGCTGCTGGATATCACACTGAGTGTGTCTC 1420

XX 61 CTACATCTCTGAGACCTTAACATACATGCTGCTGGATATCACACTGAGTGTGTCTC 120

XX 1421 AACAGTGTGACAGCAGACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1480

XX 121 AACAGTGTGACAGCAGACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 180

XX 1481 AACCTTATGTGAGCAGACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1540

XX 181 AACCTTATGTGAGCAGACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 240

XX 1541 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1600

XX 241 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 300

XX 1601 ACAGTATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTGTGCTAGAAATGCTTGTGCTG 1660

XX 301 ACAGTATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTGTGCTAGAAATGCTTGTGCTG 360

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 361 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 403

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 361 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 403

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 361 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 403

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 361 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 403

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 361 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 403

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 361 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 403

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 361 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 403

XX 30-JUL-2001; 2001WO-US23826.
 PF 03-AUG-2000; 2000US-223265P.
 XX 02-OCT-2000; 2000US-237406P.
 PR 20-MAR-2001; 2001US-277495P.
 PR 03-JUL-2001; 2001US-302702P.
 XX
 XX (CORI-) CORIXA CORP.
 PA
 PI Pyle RA, Xu J, Secretist H;
 XX
 XX WPI; 2002-257462/30.
 DR
 XX
 XX Novel polynucleotide encoding colon tumour polypeptides, useful as
 PT vaccines for treating colon cancers
 XX
 PS Claim 1; Page 206; 425bp; English.

XX The invention relates to isolated polynucleotides (I) encoding colon

XX tumour polypeptides (II). (I) is useful for stimulating an immune

XX response in a patient and treating colon cancer in a patient.

XX Oligonucleotides derived from (I) are useful for determining the presence

XX of cancer in a patient. (I) and (II) are useful in pharmaceutical

XX compositions, e.g. vaccines, and other compositions for the diagnosis

XX and treatment of colon cancer. A composition comprising a first component

XX selected from physiologically acceptable carriers and immunostimulants,

XX and an antigen-presenting cell expressing (II) is useful for inhibiting

XX development of cancer in a patient. (I) is useful in the design and

XX preparation of ribozyme molecules for inhibiting expression of tumour

XX polypeptides and (I). ABK5431-ABK5464 represent human colon cancer cDNA

XX sequences of the invention.

XX Sequence 517 BP; 121 A; 130 C; 124 G; 142 T; 0 other;

XX Query Match 19.7%; Score 403; DB 24; Length 517;

XX Best Local Similarity 100.0%; Pred. No. 2.6e-189;

XX Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

XX 1301 GGCTTGGTGCACACCTGCTGTCAGCTCAGTCACTTCTCTGAAATGCTTCTGAGCCTTGC 1360

XX 13 GGCTTGGTGCACACCTGCTGTCAGCTCAGTCACTTCTCTGAAATGCTTCTGAGCCTTGC 72

XX 1361 CTACATCTCTGAGACCTTAACATACATGCTGCTGGATATCACACTGAGTGTGTCTC 1420

XX 73 CTACATCTCTGAGACCTTAACATACATGCTGCTGGATATCACACTGAGTGTGTCTC 132

XX 1421 AACAGTGTGACAGCAGACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1480

XX 133 AACAGTGTGACAGCAGACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 192

XX 1481 AACCTTATGTGAGCAGACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1540

XX 193 AACCTTATGTGAGCAGACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 252

XX 1541 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1600

XX 253 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 312

XX 1601 ACAGTATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTGTGCTAGAAATGCTTGTGCTG 1660

XX 313 ACAGTATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTGTGCTAGAAATGCTTGTGCTG 372

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 1661 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 1703

XX 373 GATCTTCTCTGCTGGGACACATCTTGTGTCATGCTTGTGCTAGAAATGCTTGTGCTG 415

XX 08-FEB-2001 (first entry)
 XX Human ORF1711 polynucleotide sequence SEQ ID NO:3421.
 DE
 XX
 KM Human; open reading frame: ORF; detection: cytostatic; hepatotropic;
 KM vulnarary; antipariatic; antiparkinsonian; noctropic; neuroprotective;
 KM anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KM hypotensive; dermatological; immunosuppressive; antiinflammatory;
 KM antiviral; antibacterial; antifungal; antirheumatic; antihypertoid;
 KM antinaemic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KM thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR P-PSDB; AAB41947.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 2597-2599; 5507pp; English.
 XX
 CC AAC7446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORF open reading frames 1 to 3161. The ORF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnarary;
 CC antipariatic; antiparkinsonian; noctropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypertoid; and antinaemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORF-associated disorder. The
 CC nucleic acids can be used to express ORF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other;
 SO
 Query Match 15.0%; Score 306; DB 21; Length 2988;
 Best Local Similarity 100.0%; Pred. No. 3.7e-141;
 Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 452 TTGATGCCCTACATGGAACCTGGNCCCGGAGACAGTCAGCCTTTAGTGGAGAAC 511
 |||||||
 DB 1480 TTGATGCCCTACATGGAACCTGGTCCCGGAGACAGTCAGCCTTTAGTGGAGAAC 1539
 OY 512 AGCCGGGCCCCGTGTGTCGACCTGCGTGCATATCCACAGATGAATCATCCCC 571
 1540 AGCCGGGCCCCGTGTGTCGACCTGCGTGCATATCCACAGATGAATCATCCCC 1599
 DB 1540 AGCCGGGCCCCGTGTGTCGACCTGCGTGCATATCCACAGATGAATCATCCCC 1599
 OY 572 CGGGCTCAGTCAGAGCTCCCTGTCAGTCACACACCCCTTTAGGTGGTGGAGAGCCCTGC 631
 |||||||
 DB 1600 CGGGCTCAGTCAGAGCTCCCTGTCAGTCACACACCCCTTTAGGTGGTGGAGAGCCCTGC 1659
 OY 632 CGCTCTACAGCCAGCTGTCGTCAGAGAGAGTGGCTTTCAACCTGCAGTCCCTAC 691
 1660 CGCTCTACAGCCAGCTGTCGTCAGAGAGAGTGGCTTTCAACCTGCAGTCCCTAC 1719
 DB 1660 CGCTCTACAGCCAGCTGTCGTCAGAGAGAGTGGCTTTCAACCTGCAGTCCCTAC 1719
 OY 692 CGCCTGCTGAAGAACCCCTCCCTCAACCTGCATATGTCGACCTGTCGGGACCCCGG 751
 |||||||
 DB 1720 CGCCTGCTGAAGAACCCCTCCCTCAACCTGCATATGTCGACCTGTCGGGACCCCGG 1779
 OY 752 GCCGTG 757
 |||||||
 DB 1780 GCCGTG 1785
 RESULT 10
 ABR54794
 ID ABR54794 standard; cDNA; 389 BP.
 XX
 AC ABR54794;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Human colon cancer-associated cDNA, SEQ ID NO 264.
 XX
 KM Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200212280-A2.
 XX
 PD 14-FEB-2002.
 XX
 PF 30-JUL-2001; 2001MO-US23826.
 XX
 PR 03-AUG-2000; 2000US-223265P.
 PR 02-OCT-2000; 2000US-237406P.
 PR 20-MAR-2001; 2001US-277495P.
 PR 03-JUL-2001; 2001US-302702P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Pyle RA, Xu J, Secrist H;
 XX
 DR WPI; 2002-257462/30.
 XX
 PT Novel polynucleotide encoding colon tumour polypeptides, useful as
 PT vaccines for treating colon cancers -
 XX
 PS Claim 1; Page 225; 425pp; English.
 XX
 CC The invention relates to isolated polynucleotides (I) encoding colon
 CC tumour polypeptides (II). (I) is useful for stimulating an immune
 CC response in a patient and treating colon cancer in a patient.
 CC Oligonucleotides derived from (I) are useful for determining the presence
 CC of cancer in a patient. (II) and (II) are useful in pharmaceutical
 CC compositions, e.g. vaccines, and other compositions for the diagnosis
 CC and treatment of colon cancer. A composition comprising a first component
 CC selected from physiologically acceptable carriers and immunostimulants,
 CC and an antigen-presenting cell expressing (II) is useful for inhibiting
 CC development of cancer in a patient. (I) is useful in the design and
 CC preparation of ribozyme molecules for inhibiting expression of tumour

CC polypeptides and (I). ABR54531-ABR55464 represent human colon cancer cDNA
 CC sequences of the invention.
 XX
 SQ Sequence 389 BP; 121 A; 83 C; 76 G; 109 T; 0 other;

Query Match 5.6%; Score 114; DB 24; Length 389;
 Best Local Similarity 100.0%; Pred. No. 6.9e-46;
 Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1930 TGGCATTCTGTAACTAAATTCACAAATTAAGTCTGTAGATGTCCTTTTATG 1989
 DB 266 TGGCATTCTGTAACTAAATTCACAAATTAAGTCTGTAGATGTCCTTTTATG 325

OY 1990 CTTCTTAATTATTAGCAGTAAATGTTTCATTTTATGGATGCTTAA 2043
 DB 326 CTTCTTAATTATTAGCAGTAAATGTTTCATTTTATGGATGCTTAA 379

RESULT 11

AAK92588/c
 ID AAK92588 standard; cDNA; 548 BP.

AC AAK92588;

DT 06-NOV-2001 (first entry)

DE Human cDNA 3'-end sequence, SEQ ID NO: 1048.

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

OS Homo sapiens.

PN EPI130094-A2.

PD 05-SEP-2001.

PF 07-JUL-2000; 2000EP-0114089.

PR 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI: 2001-524255/58.

DR 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -

PS Claim 3; SEQ ID NO 1048; 1380pp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesized by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is the nucleotide
 CC sequence of the 3'-end of a cDNA provided in the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 548 BP; 160 A; 109 C; 129 G; 148 T; 2 other;

Query Match 5.1%; Score 105; DB 22; Length 548;
 Best Local Similarity 100.0%; Pred. No. 2e-41;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1930 TGGCATTCTGTAACTAAATTCACAAATTAAGTCTGTAGATGTCCTTTTATG 1989
 DB 108 TGGCATTCTGTAACTAAATTCACAAATTAAGTCTGTAGATGTCCTTTTATG 49

OY 1990 CTTCTTAATTATTAGCAGTAAATGTTTCATTTTATGGATGCTTAA 2034
 DB 48 CTTCTTAATTATTAGCAGTAAATGTTTCATTTTATGGATGCTTAA 4

RESULT 12

AAAD24670
 ID AAD24670 standard; cDNA; 1647 BP.

AC AAD24670;

DT 12-MAR-2002 (first entry)

DE Human drug metabolizing enzyme (DME)-5 cDNA.

XX Human; drug metabolizing enzyme; gene therapy; autoimmune disorder;
 KW inflammatory disorder; acquired immune deficiency syndrome; infection;
 KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
 KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
 KW cancer; endocrine disorder; hypothalamus disorder; pituitary disorder;
 KW gastrointestinal disorder; metabolic disorder; developmental disorder;
 KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
 KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
 KW DME-5; ss.

XX Homo sapiens.

XX Key Location/Qualifiers

EH CDS 1..1188

FT /tag- a /product- "Human DME-5 protein"

FT sig-peptide 1..96

FT mat-peptide 97..1185

FT /tag- c /product- "Human mature DME-5 protein #1"

FT sig-peptide 1..105

FT mat-peptide 106..1185

FT /tag- d /product- "Human mature DME-5 protein #2"

FT mat-peptide 106..1185

FT /tag- e /product- "Human mature DME-5 protein #2"

FT mat-peptide 106..1185

FT /tag- f /product- "Human mature DME-5 protein #2"

FT mat-peptide 106..1185

FT /tag- g /product- "Human mature DME-5 protein #2"

FT mat-peptide 106..1185

FT /tag- h /product- "Human mature DME-5 protein #2"

FT mat-peptide 106..1185

FT /tag- i /product- "Human mature DME-5 protein #2"

FT mat-peptide 106..1185

FT /tag- j /product- "Human mature DME-5 protein #2"

FT mat-peptide 106..1185

FT /tag- k /product- "Human mature DME-5 protein #2"

FT mat-peptide 106..1185

FT /tag- l /product- "Human mature DME-5 protein #2"

PS Claim 5; Page 139; 143pp; English.

XX The invention relates to human drug metabolizing enzymes referred as
 CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides
 CC of the invention are useful for assessing toxicity of test compounds
 CC and in gene therapy. Sequences of the invention are useful in the
 CC diagnosis, prevention and treatment of autoimmune/inflammatory
 CC disorders such as acquired immune deficiency syndrome (AIDS), adult
 CC respiratory distress syndrome, allergies, anaemia, atherosclerosis,
 CC asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's
 CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's
 CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple
 CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus
 CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,
 CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic
 CC infections; cell proliferative disorders such as actinic keratosis,
 CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,
 CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,
 CC epilepsy; endocrine disorders such as disorders of the hypothalamus
 CC and pituitary resulting from lesions such as primary brain tumours,
 CC adenomas, infarction associated with pregnancy, aneurysms, vascular
 CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,
 CC glaucoma, pigmentosa; metabolic disorders such as Addison's disease,
 CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,
 CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,
 CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such
 CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,
 CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
 CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
 CC disorders. The present sequence is human DME-5 protein cDNA.

XX
 SQ Sequence 1647 BP; 287 A; 560 C; 501 G; 299 T; 0 other;

Query Match 2.9%; Score 59; DB 24; Length 1647;
 Best Local Similarity 100.0%; Pred. No. 1.4e-18;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCGCTCTACAGCAGCGTGCTCAAGAGGCGCTTCTTCAACCTGCAG 682
 |||||||
 DB 488 AGGCTGCGCGCTCTCTACAGCAGCGTGCTCAAGAGGCGCTTCTTCAACCTGCAG 546

RESULT 13
 AAD02700
 ID AAD02700 standard; cDNA; 1694 BP.

XX
 AC AAD02700;
 XX
 DT 02-MAY-2001 (first entry)

XX
 DE Human glycosyl sulfotransferase-4beta (GST-4beta) cDNA.

XX
 KW Human: glycosyl sulfotransferase-4beta; GST-4beta; Immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenailitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1; ss.

XX
 OS Homo sapiens.

XX
 FH Key Location/Qualifiers
 FT 5'UTR 9..188
 FT /*tag- a
 FT CDS 189..1376
 FT /*tag- b
 FT /*product= "Human glycosyl sulfotransferase-4beta
 FT (GST-4beta)"
 FT /note= "CDS is referred as SEQ ID NO:21 in brief

FT description of the figures (page no: 4)"
 FT 3'UTR 1377..1694
 FT /*tag- c

XX
 XX WO200106015-A1.
 XX
 XX 25-JAN-2001.
 XX
 XX 19-JUL-2000; 2000WO-0519741.
 XX
 XX 20-JUL-1999; 9905-0144694.
 XX PR 13-JUL-2000; 2000US-0593828.
 XX
 XX (REGC) UNIV CALIFORNIA.
 XX
 XX Rosen SD, Lee JK, Hemmerich S;
 XX WPI; 2001-138471/14.
 XX P-PSDB; AAY72640.

DR
 DR P-PSDB; AAY72640.

PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications -

PS Claim 6; Fig 4A; 128pp; English.

XX
 XX The present sequence is human glycosyl sulfotransferase-4beta (GST-4
 XX beta) cDNA. GST-4 gene is found on chromosome 16q23.1.
 XX GST is a type 2 membrane protein useful for inhibiting a binding event
 XX between a selectin and a selectin ligand, which comprises contacting the
 XX selectin with a non-sulphated selectin ligand, GST and a small molecular
 XX agent that inhibits the sulphation activity of GST. GST is also useful
 XX in inhibiting a selectin mediated binding event. GST is useful in gene
 XX therapy to treat disorders such as acute or chronic inflammation,
 XX systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 XX nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 XX disease, Grave's disease, adrenalitis, hypoparathyroidism, periculous
 XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
 XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 XX during transplantation.

XX
 SQ Sequence 1694 BP; 259 A; 586 C; 514 G; 335 T; 0 other;

Query Match 2.9%; Score 59; DB 22; Length 1694;
 Best Local Similarity 100.0%; Pred. No. 1.4e-18;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCGCTCTACAGCAGCGTGCTCAAGAGGCGCGCTTCTTCAACCTGCAG 682
 |||||||
 DB 676 AGGCTGCGCGCTCTCTACAGCAGCGTGCTCAAGAGGCGCGCTTCTTCAACCTGCAG 734

RESULT 14
 AAD02699
 ID AAD02699 standard; cDNA; 2044 BP.

XX
 AC AAD02699;
 XX
 DT 02-MAY-2001 (first entry)

XX
 DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.

XX
 KW Human: glycosyl sulfotransferase-4alpha; GST-4alpha; Immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adenailitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;

Fri Aug 15 08:09:29 2003

us-09-645-078-1_1.rng

Page 13

Db 834 AGGCTGCGCCTCTCTACAGCCACGTGGTGCTCAAGAGAGTGGCGTTCTTCAACCTGCAG 892

Search completed: August 15, 2003, 01:05:21
Job time : 546 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2003, 00:44:22 ; Search time 3887 Seconds
(without alignments)
12774.388 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043
Sequence: 1 gaattccattgtgtgtgta.....tgggacccaataaaaaa 2043

Scoring table: OLIGO_NMC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45454738

Minimum DB seq length: 50

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estbda:*
2: em_estbhum:*
3: em_estlin:*
4: em_estlinu:*
5: em_estlov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_estl:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	685	33.5	954	BI823850 603039012
2	579	28.3	583	BK490456 Homo sapi
3	534	26.1	668	AL709927 DKFZP6860
4	479	23.4	571	BM129080 1F17C04.Y

C	5	322	15.8	593	12	BM969292	BM969292 UI-CF-ENO
C	6	290	14.2	680	10	BF878439	BF878439 MRO-ET010
C	7	288	14.1	669	13	B0684395	B0684395 UI-CF-ENO
C	8	218	10.7	553	12	BM128370	BM128370 1F13d12.x
C	9	218	10.7	553	12	BM128370	BM128370 1F17C04.x
C	10	107	5.2	362	9	AM002418	AM002418 wu61903.x
C	11	105	5.1	417	9	AM572510	AM572510 xq18911.x
C	12	101	4.9	358	10	BF056840	BF056840 7K10d12.x
C	13	74	3.6	419	9	AM572390	AM572390 xu09a06.x
C	14	59	2.9	525	10	BF197521	BF197521 7084d08.x
C	15	59	2.9	620	9	AL824100	AL824100 w146c01.x
C	16	59	2.9	695	10	BE857538	BE857538 7901a08.x
C	17	59	2.9	735	10	BE858652	BE858652 7901a09.x
C	18	55	2.7	721	28	A0741911	A0741911 HS-5569.B
C	19	47	2.3	436	9	AM081348	AM081348 xc41b06.x
C	20	27	1.3	507	12	BM246681	BM246681 K0741E04-
C	21	27	1.3	571	12	BM245312	BM245312 K0722H07-
C	22	27	1.3	695	12	BG964671	BG964671 602831875
C	23	27	1.3	783	12	BG963298	BG963298 602827716
C	24	27	1.3	852	12	BG966340	BG966340 602832826
C	25	27	1.3	1923	11	AK009113	AK009113 Mus muscu
C	26	23	1.1	362	10	BE340900	BE340900 EST344988
C	27	23	1.1	397	9	AL345721	AL345721 CBd4d04.x
C	28	23	1.1	687	14	CB167190	CB167190 GNM602702
C	29	23	1.1	692	14	BY711808	BY711808 BY711808
C	30	23	1.1	834	13	B0884877	B0884877 ACENCOUPT
C	31	22	1.1	148	9	AL345073	AL345073 lb62b07.x
C	32	22	1.1	160	13	BQ255877	BQ255877 NISC_f09
C	33	22	1.1	167	9	AL253971	AL253971 qv34a05.x
C	34	22	1.1	167	9	AL733948	AL733948 qv54a05.x
C	35	22	1.1	171	9	AL246069	AL246069 qv52f03.x
C	36	22	1.1	187	9	AL251238	AL251238 qv55g05.x
C	37	22	1.1	187	9	AL792899	AL792899 qv55g05.y
C	38	22	1.1	191	9	AL251979	AL251979 qv57d08.x
C	39	22	1.1	196	9	AL792857	AL792857 qv34a05.y
C	40	22	1.1	253	9	AA076760	AA076760 7B01F06 C
C	41	22	1.1	290	9	AL345395	AL345395 lb82b03.x
C	42	22	1.1	361	9	AL311365	AL311365 ta47h05.x
C	43	22	1.1	493	10	BF869376	BF869376 IL3-EP011
C	44	22	1.1	501	28	BH287165	BH287165 CH230-113
C	45	22	1.1	516	9	AA689056	AA689056 vs06c06.x

ALIGNMENTS

RESULT 1
BI823850
LOCUS
DEFINITION
603039012F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179826 5',
mRNA sequence.
BI823850
VERSION
BI823850.1 GI:15935400
KEYWORDS
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
REFERENCE
NIH-MGC http://mgc.nci.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
CONTACT: Robert Strausberg, Ph.D.
Email: gcapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
plate: L1AM11448 row: h column: 03
High quality sequence stop: 856.

OY	281	GGCTGTGGCTCTCTTTTGTGGGAGCTTTTGGGAGACCCAGATGTTTCTACTG	340
Db	241	CGCTGTGGCTCTCTTTGTGGGAGCTTTTGGGAGACCCAGATGTTTCTACTG	300
OY	341	ATGGAGCCGCGCTGTGGACGTGTGATGACCTTCAAGCAGACACCGCTGTGATGCTGCAC	400
Db	301	ATGGAGCCGCGCGCTGTGGACGTGTGATGACCTTCAAGCAGACACCGCTGTGATGCTGCAC	360
OY	401	ATGGCTGTGGGGATCTGATACGGGGCGCTCTTCTTGTGCGACATGAGCGTCTTGTGATCC	460
Db	361	ATGGCTGTGGGGATCTGATACGGGGCGCTCTTCTTGTGCGACATGAGCGTCTTGTGATCC	420
OY	461	TACATGGAACCTGCTGCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAGAACAGCCGGGCC	520
Db	421	TACATGGAACCTGCTGCCCGGAGAGAGTCCAGCCTCTTTCAGTGGGAGAACAGCCGGGCC	480
OY	521	CTGTGTTCTGACCGCTGTGACATCATATCCCAACAAAGATGAAATCATCCCCGGGCTCAC	580
Db	481	CTGTGTTCTGACCGCTGTGACATCATATCCCAACAAAGTGAATCATCCCCGGGCTCAC	540
OY	581	TGCAGGCTCTGTGACGTCAACACAGCCCTTGTAGGAGTG	619
Db	541	TGCAGGCTCTGTGACGTCAACACAGCCCTTGTAGGAGTG	579

RESULT 3	AL709927	LOCUS	DEFINITION
	AL709927	668 bp	mRNA
	DKE2p68602364-r1	686 (synonym: hlcc3)	linear
	DKE2p68602364 5'		Homo sapiens cDNA clone
	126002		mRNA sequence.

ACCESSION	AL/0992/	
VERSION	AL709927.1	GI:19693282

SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE 1 (bases 1 to 668)

AUTHORS	TITLE
Bahr, A., Lauber, J., Mewes, H. W., Well, B., Amlid, C., Osanger, A., Fobbe, G., Han, M. and Wiemann, S.	EST (Bahr, A., Lauber, J., Mewes, H. W., Well, B., et al.)

JOURNAL Unpublished
COMMENT Contact: Bahr A
MIPS

Ingolstraedter Landstr.1, D-85764 Neuberg, Germany
This is the 5' sequence of the clone insert
Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
sequenced by Jägen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.

This clone (DKFZp68602364) is available at the RZPD in Berlin. Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES
Source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp686O2364"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/clone_lib="686 (synonym: hlyc3)"
/notes="Vector: pTR1plex2; site_1: sf11A; site_2: sf11B
cDNA-collection"
BASE COUNT      132 a      202 c      178 g      154 t      2 others
ORIGIN

```

Query Match	26.1%	Score 534	DB 9	Length 668
Best Local Similarity	100.0%	Pred. No. 2.3e-209		
Matches 534	Conservative 0	Mismatches 0	Indels 0	Gaps 0

QY	40	AGAGGAGAAAAGCGCATGGCCCGCGGTAGAGAGGACGCTCCAAAAGCACAGGAAAGCCC	99
Db	6	AGAGGAGAAAAGCGCATGGCCCGCGGTAGAGAGGACGCTCCAAAAGCACAGGAAAGCCC	65
QY	100	AAGGCACAAAGGCTCTCCACTTCAGCACATGCTACTGCTAAAAAAAATAGACTCCTGCT	159
Db	66	AAGGCACAAAGGCTCTCCACTTCAGCACATGCTACTGCTAAAAAAAATAGACTCCTGCT	125
QY	160	GTTTCGTGTTTCCAGATGGCCATCTTGCTCTATTCTTCACATGTATACGACACAAAT	219
Db	126	GTTTCGTGTTTCCAGATGGCCATCTTGCTCTATTCTTCACATGTATACGACACAAAT	185
QY	220	CAGCTCCCTGTCTATGAAGGACACAGCCGAGCGCATGCACGTGCTGTCTCTCTG	279
Db	186	CAGCTCCCTGTCTATGAAGGACACAGCCGAGCGCATGCACGTGCTGTCTCTCTG	245
QY	280	GCGCTCTGGCTCTCTTTGTGTGGGACGTTTTGTGGGACGACCCGATGTTTCTACCT	339
Db	246	GCGCTCTGGCTCTCTTTGTGTGGGACGTTTTGTGGGACGACCCGATGTTTCTACCT	305
QY	340	GATGAGAGCCCGCTGGACAGTGTGATGACTTTCAAGCAGACACCGCTGTGATCTGCA	399
Db	306	GATGAGAGCCCGCTGGACAGTGTGATGACTTTCAAGCAGACACCGCTGTGATCTGCA	365
QY	400	CATGGCTGTGGGGATCTGATACGGGGCGCTCTTCTGTGTCGACATGAGCGCTTTTGATGC	459
Db	366	CATGGCTGTGGGGATCTGATACGGGGCGCTCTTCTGTGTCGACATGAGCGCTTTTGATGC	425
QY	460	CTACATGAACTGGTCCCCCGAGACAGTCCAGGCTCTTTCAGTGGGAAACAGCGGGC	519
Db	426	CTACATGAACTGGTCCCCCGAGACAGTCCAGGCTCTTTCAGTGGGAAACAGCGGGC	485
QY	520	CCTGATGTTCTGCACTGGCTGTGATATATCCACAAAGATGAAATCATGCCCG	573
Db	486	CCTGATGTTCTGCACTGGCTGTGATATATCCACAAAGATGAAATCATGCCCG	539

RESULT 4

BM129080	LOCUS	DEFINITION	571 bp	mRNA	linear	EST 12-MAR-200
BM129080	BM129080	Melton Normalized Human Islet 4 N4-HS 1 Homo sapiens				

CDNA CLONE IMAGE:5676655 5' similar to TR:Q9Y5R3 Q9Y5R3
N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE. ; mRNA sequence.
BM129080

VERSION BM129080.1 GI:17123632
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

REFERENCE
AUTHORS
1 (bases 1 to 5/1)
Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
Lemshaha, I., Searce, M., Brestell, J., Gadowh, G., Clifton, S.,
Hillier, L., Marra, M., Page, D., Wylie, T., Martin, J., Blstein, A.,

Schmitt, A., Theising, B., Rittler, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarelshvili, R., Williams, T

TITLE
JOURNAL
COMMENT

Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished
Other-ESTs: 1f17c04.x1
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@bioph.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center For information on obtaining a clone please contact: Juliana Brown (brownjefas.harvard.edu) This sequence now available from the IMAGE consortium, for clone orders contact: info@image.llnl.gov

High quality sequence stop: 434

FEATURES	Location/Qualifiers
source	1. .571

```

/organism="Homo sapiens"
/mol_type="mRNA"

```

```
/db_xref="taxon:9606"  
/clone="IMAGE:5676655"
```

/sex="Both"

/tissue_type="Islets of Langerhans"

```
/dev_stage="Adult"
```

```
/usr/lib/perl5/vendor_perl/5.10.1/CORE/libperl.so
```

/note="Organ: Pancreas; Vector: pSPORT1; Site 1: Not

```
Site_2: Sal 1; Starting library constructed using
```

SuperScript[®] plasmid library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation, average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldi, Lennon, and Soares 1996 Genome Research 6:791-806. 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

KEYWORDS	EST.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens

REFERENCE
1 (bases 1 to 593)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
JOURNAL discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477

PUBMED 8889548
COMMENT Contact: McCray, PB

McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA-Yes.

FEATURES	Location/Qualifiers
source	1. .593

```
/organism="Homo sapiens"
/mol_type="mRNA"
```

```

/db_xref="taxon:9606"
/clone="UI-CF-ENDO-acp-1-21-0-UI"

```

/tissue_type="Primary Lung Cystic Fibrosis Epithelial Cells"

```
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
```

```

/clone_id="OI-CF-EN0"
/note="Organ: Lung; Vector: pT7T3-Pac (Pharmacia) with a

```

modified polylinker; site_1: EcoR I; site_2: Not I; UI-CF-ENO is a cDNA library containing the following

tissue(s): Primary Lung Cystic Fibrosis Epithelial cells. The library was constructed according to Bonaldo, Lennon

CDNA synthesis was primed with an oligo-dT primer

containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into pT7T3-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)₁₈ tail. The sequence tag for this library is CTGCTCAGGT.

TAG_LIB-U1-CF-ENO
TAG_TISSUE-Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG_SEO-CTGCTCAGGP"

|||||
 Db 473 CCCACCGTGAACAGGGGTATGCTCTTCTTTCTTGTATCTTCTGCTGCTGAGAC 414
 Oy 1562 TTTCGAGACTTTGGGCTGAGGCTTATTAAGACACACAGATATAGTGATATC 1621
 Db 413 TTTCGAGACTTTGGGCTGAGGCTTATTAAGACACAGATATAGTGATATC 354
 Oy 1622 CATTAACCTCCCTGTCACATCTTGGCCAAATGGGATGATCTTTCACCAAGAAGCTCA 1681
 Db 353 CATTAACCTCCCTGTCACATCTTGGCCAAATGGGATGATCTTTCACCAAGAAGCTCA 294
 Oy 1682 CCAGCATTTTCCACAGAGATGC 1703
 Db 293 CCAGCATTTTCCACAGAGATGC 272

RESULT 6
 BF878439
 LOCUS BF878439 680 bp mRNA linear EST 17-JAN-2001
 DEFINITION MRO-ET0109-191100-002-h06 ET0109 Homo sapiens cDNA, mRNA sequence.
 ACCESSION BF878439
 VERSION BF878439.1 GI:12268569
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 680)
 Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M.R., Nagai, M.A., da Silva, W. Jr., Zago, M.A., Bordu, S., Costa, F.F., Goldman, G.H., Carvalho, A.F., Matsukuma, A., Bala, G.S., Simpson, D.H., Brunstein, A., de Oliveira, P.S., Bucher, P., Jongeneel, C.V., O'Hare, M.J., Soares, F., Brentani, R.R., Reis, L.F., de Souza, S.J. and Simpson, A.J.
 Shotgun sequencing of the human transcriptome with ORF expressed sequence tags
 Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
 20202663
 PUBMED 10737800
 COMMENT Contact: Simpson A.J.G.
 Laboratory of Cancer Genetics
 Ludwig Institute for Cancer Research
 Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP, Brazil
 Tel: +55-11-2704922
 Fax: +55-11-2707001
 Email: asimpson@ludwig.org.br
 This sequence was derived from the PAPSP/LICR Human Cancer Genome Project. This entry can be seen in the following URL
 (http://www.ludwig.org.br/scripts/gethtml2.pl?tl-MRO<2-MRO-ET0109-191100-002-h06<3-2000-11-19<4-1)
 Seq primer: puc 18 forward
 High quality sequence start: 17
 High quality sequence stop: 641.
 Location/Qualifiers
 1. 680
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /dev_stage="adult"
 /clone_1lb="ET0109"
 /note="Organ: lung; tumor; Vector: puc18; site_1: Sma1; site_2: Sma1; A mini-library was made by cloning products derived from ORESTES PCR (U.S. Letters Patent application No. 196,716 - Ludwig Institute for Cancer Research) profiles into the puc 18 vector. Reverse transcription of tissue mRNA and cDNA amplification were performed under low stringency conditions."
 BASE COUNT 168 a 164 c 163 g 184 t 1 others
 ORIGIN
 Query Match 14.2%; Score 290; DB 10; Length 680;
 Best Local Similarity 100.0%; Pred. No. 9.9e-109;

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Oy 1414 TGTGTCCACACGTCCTCAGAGAGACTTTTGTGTCATGCTGTGTAGAAACAG 1473
 Db 106 TGTGTCCACACGTCCTCAGAGAGACTTTTGTGTCATGCTGTGTAGAAACAG 165
 Oy 1474 ACTGGGAACTTATGTGACAGACATCCACCAGGAAACAGGATATGCTCTTC 1533
 Db 166 ACTGGGAACTTATGTGACAGACATCCACCAGGAAACAGGATATGCTCTTC 225
 Oy 1534 TTTTCTTGATCTTCTGTCGCGAGACTTCAGAGACTTGTGCGCTGAGGCTATTAA 1593
 Db 226 TTTTCTTGATCTTCTGTCGCGAGACTTCAGAGACTTGTGCGCTGAGGCTATTAA 285
 Oy 1594 GCACGACACAGTATCATGTGAATGATCCATTAACCTCCCTGTGCTGATGCTCCCAATG 1653
 Db 286 GCACGACACAGTATCATGTGAATGATCCATTAACCTCCCTGTGCTGATGCTCCCAATG 345
 Oy 1654 GGGATGATCTTTCACCAAGAGCTCAGACATTTTCCACAGAGATGC 1703
 Db 346 GGGATGATCTTTCACCAAGAGCTCAGACATTTTCCACAGAGATGC 395

RESULT 7
 BU684395/c
 LOCUS BU684395 669 bp mRNA linear EST 07-OCT-2002
 DEFINITION UI-CF-ENO-aco-f-08-0-UI-s1 UI-CF-ENO Homo sapiens cDNA clone
 UI-CF-ENO-aco-f-08-0-UI 3', mRNA sequence.
 ACCESSION BU684395
 VERSION BU684395.1 GI:23537302
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 669)
 Bonaldo, M.F., Lennon, G. and Soares, M.B.
 Normalization and subtraction: two approaches to facilitate gene discovery
 Genome Res. 6 (9), 791-806 (1996)
 9704477
 PUBMED 8889548
 COMMENT Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul.mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 cDNA library preparation: Dr. M. Bento Soares, University of Iowa
 DNA library Arrayed by: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research Genetics (www.resgen.com) or from Open Biosystems (www.openbiosystems.com).
 Seq primer: M13 FORWARD
 Polya=Yes.
 Location/Qualifiers
 1. 669
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone_1lb="UI-CF-ENO"
 /tissue="Primary Lung Cystic Fibrosis Epithelial Cells"
 /dev_stage="adult"
 /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
 /clone_1lb="UI-CF-ENO"
 /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified polylinker; Site_1: EcoR I; Site_2: Not I; UI-CF-ENO is a cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells."

```

obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.lnml.gov
High quality sequence stop: 425.
Location/Qualifiers
1..553
source
```

amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806: 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library.

BASE COUNT	163 a	111 c	127 g	152 t
ORIGIN				

Best Local Similarity 100.0%; Pred. No. 5.3e-79;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0

Db 553 TANGTAGAGACACATCCACCACAGTGAAACAGGGATTTGCTCTTCTTTTCTTGATCT 494

Db 493 TCCTGTCGTGGCAGACTTCAGAGACTTGTGTGGCCGTATTTAAACACGACAGACT 434

Dy 1606 ATCAGTGGGAATTGATCCATTAACCTCCCTGTGCACATCTTGCCCAATGGGGAATGATCT 166

433 ATTAGTGGGAATTGATCCATTAACCTCCCTGTGCACATCTTGCCCAATGGGGAATGATCT 374

QY	1666	TTTCACCAAGAGCTTCACGACATTTTCCACAGAGATGC	1703
Db	373	TTTCACCAAGAGCTTCACGACATTTTCCACAGAGATGC	336

RESULT 9	553 bp	mrna	linear	EST 12-MAR-2000
BM128831/c				
LOCUS				
BM128831				

ALGORITHM	SOURCE	ORGANISM
EST	Homo sapiens (human)	
	Homo sapiens	
	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;	

TITLE Hillier, L., Marra, M., Page, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarashvili, R., Williams, T., Jackson, Y., and Bowers, Y.

Endocrine Pancreas Consortium

JOURNAL
COMMENT

Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: Juliana Brown
(brownjefas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 425.

FEATURES

SOURCE

Location/Qualifiers
1..553

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5676655"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/note="Organ: Pancreas; Vector: pSPORT1; Site:1: Not 1;
Site:2: Sal 1; Starting library constructed using
SuperScript Plasmid Library Kit (Life Technologies). cDNA
made by oligo-dt priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 163 a 111 c 127 g 152 t
ORIGIN

Query Match 10.7%; Score 218; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.3e-79;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1486 TATGTAGCAGACATCCACAGGATGATGCTCTTCTTTCTTGATCT 1545
DB 553 TATGTAGCAGACATCCACAGGATGATGCTCTTCTTTCTTGATCT 494
QY 1546 TCCGTGTGGGACACTTCAGAGACTTGGGCTGAGGCTATTAAAGCAGCAGT 1605
DB 493 TCCGTGTGGGACACTTCAGAGACTTGGGCTGAGGCTATTAAAGCAGCAGT 434
QY 1606 ATCAGTGAATTTGATCCATAAACCCTCCCTGTCACATCTTGCCCAATGGGGAAGGATCT 1665
DB 433 ATCAGTGAATTTGATCCATAAACCCTCCCTGTCACATCTTGCCCAATGGGGAAGGATCT 374
QY 1666 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1703
DB 373 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 336

RESULT 10
AM002418/c 362 bp mRNA linear EST 27-OCT-1999
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
AUTHORS
TITLE
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL
COMMENT

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.
cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LINL at:
www-bio.llnl.gov/bdrp/image/image.html
Seq primer: -400p from gibco.
Location/Qualifiers
1..362

FEATURES

SOURCE

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2524564"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/clone_id="NCI-CGAP-GC6"
/note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site:1: Not 1; Site:2: Eco RI; Plasmid DNA
from the normalized library NCI-CGAP-GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 108 a 69 c 77 g 108 t
ORIGIN

Query Match 5.2%; Score 107; DB 9; Length 362;
Best Local Similarity 100.0%; Pred. No. 3.7e-33;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGGCATCTGTTAACTAAATTCACAAATGATGCTTTAGAAATGCCCTTTATG 1989
DB 107 TGGCATCTGTTAACTAAATTCACAAATGATGCTTTAGAAATGCCCTTTATG 48
QY 1990 CTCTTAAATTAATTTAGAGTAATGTCAATTTTATGGATCTTAA 2036
DB 47 CTCTTAAATTAATTTAGAGTAATGTCAATTTTATGGATCTTAA 1

RESULT 11
AM572510/c 417 bp mRNA linear EST 13-MAR-2000
LOCUS
DEFINITION
ACCESSION
VERSION
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
AUTHORS
TITLE
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Cloning Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: image.llnl.gov/image/html/iresources.shtml
 Seq primer: -40UP from Gibco
 High quality sequence stop: 416.
 Location/Qualifiers
 1.417
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2751044"
 /tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"
 /lab_host="DH10B"
 /clone_1lb="NCI-CGAP_011"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

BASE COUNT 140 a 65 c 80 g 132 t
 ORIGIN

Query Match 5.1%; Score 105; DB 9; Length 417;
 Best Local Similarity 100.0%; Pred. No. 2.3e-32;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TCCATCTGTATTAATAAATCCCAATTAAGTCTGTTAGAAATGCCCTTTTANG 1989
 DB 266 TCCCATCTGTATTAATAAATCCCAATTAAGTCTGTTAGAAATGCCCTTTTANG 207
 QY 1990 CTCTTAATTATAGAGTAATGTTCAATTTTATGGATCCGTA 2034
 DB 206 CTCTTAATTATAGAGTAATGTTCAATTTTATGGATCCGTA 162

RESULT 12 358 bp mRNA linear EST 16-OCT-2000
 BF056840/c 7k10d12.x1 NCI-CGAP-06 Homo sapiens CDNA clone IMAGE:3443783 3',
 LOCUS mRNA sequence.
 BF056840
 ACCESSION BF056840.1 GI:10810736
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 358)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Cloning Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILN, send email to: inf@image.llnl.gov
 Seq primer: -40UP from Gibco.

FEATURES
 source
 1.358
 Location/Qualifiers
 /organism="Homo sapiens"
 /mol_type="mRNA"

/db_xref="taxon:9606"
 /clone="IMAGE:3443783"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /clone_1lb="NCI-CGAP-06"
 /note="Vector: pT7T3D-Pac (Pharmacia) with a modified polylinker; Site: 1: Not I; Site: 2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP-04 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 107 a 68 c 82 g 101 t
 ORIGIN

Query Match 4.9%; Score 101; DB 10; Length 358;
 Best Local Similarity 100.0%; Pred. No. 1.1e-30;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1932 CCATCGTAATACATAAATCCCAATTAAGTCTGTTAGAAATGCCCTTTATGCT 1991
 DB 101 CCATCGTAATACATAAATCCCAATTAAGTCTGTTAGAAATGCCCTTTATGCT 42
 QY 1992 TCTTAATTATAGAGTAATGTTCAATTTTATGGATCCCT 2032
 DB 41 TCTTAATTATAGAGTAATGTTCAATTTTATGGATCCCT 1

RESULT 13 419 bp mRNA linear EST 13-MAR-2000
 AM572390/c XU09a06.x2 NCI-CGAP-014 Homo sapiens CDNA clone IMAGE:2799634 3',
 LOCUS mRNA sequence.
 AM572390
 ACCESSION AM572390.1 GI:7237123
 VERSION EST.
 KEYWORDS Homo sapiens (human)
 SOURCE Homo sapiens
 ORGANISM Homo sapiens; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 419)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgabs-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Cloning Distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/ILN at: image.llnl.gov/image/html/iresources.shtml
 Seq primer: -40UP from Gibco
 High quality sequence stop: 343.
 Location/Qualifiers
 1.419
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2799634"
 /tissue_type="moderately-differentiated adenocarcinoma"
 /lab_host="DH10B"
 /clone_1lb="NCI-CGAP-014"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site: 1: SalI; Site: 2: NotI; Cloned unidirectionally. Primer: Oligo dT. Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"

BASE COUNT 122 a 78 c 89 g 130 t

ORIGIN

Query Match 3.6%; Score 74; DB 9; Length 419;
 Best Local Similarity 100.0%; Pred. NO. 1.4e-19;
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1630 TCCCTGTCACATCTTGCACATGGGAAATGATCTTTACCAAGAAGCTCACCAGATT 1689
 |||||||
 DB 419 TCCCTGTCACATCTTGCACATGGGAAATGATCTTTACCAAGAAGCTCACCAGATT 360
 |||||||

QY 1690 TTCACAGAGATGC 1703
 |||||||

DB 359 TTCACAGAGATGC 346

RESULT 14
 Bf197521/c 525 bp mRNA linear EST 03-NOV-2000
 LOCUS Bf197521/c 525 bp mRNA linear EST 03-NOV-2000
 DEFINITION 7684a08.x1 NCI-CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3642903 3'
 similar to TR:Q9UBY3 Q9UBY3 N-ACETYLGLUCOSAMINE
 6-O-SULFOTRANSFERASE. ; mRNA sequence.

ACCESSION Bf197521
 VERSION Bf197521.1 GI:11086670
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 525)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL, send email to:
 lnl@image.lnl.gov
 High quality sequence stop: 451.

FEATURES
 source Location/Qualifiers
 1..525
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3642903"
 /lab_host="DH10B"
 /clone_lib="NCI-CGAP_Kid11"
 /note="Organ: kidney; Vector: p773D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Plasmid DNA from the normalized library NCI-CGAP_Kid3 was
 prepared, and ss circles were made in vitro. Following HAP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library
 (cloneids 1322376-1323911, 1456007-1456775, and
 1500552-1502855). Subtraction by Bento Soares and M.
 Fatima Bonaldo."

BASE COUNT 90 a 163 c 187 g 85 t

ORIGIN

Query Match 2.9%; Score 59; DB 10; Length 525;
 Best Local Similarity 100.0%; Pred. No. 2e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCGCTCTTACAGCAGCTGCTGCTCAAGAGGCGCTTTCAACCTGCAG 682
 |||||||
 DB 260 AGGCTGCGCGCTCTTACAGCAGCTGCTGCTCAAGAGGCGCTTTCAACCTGCAG 202

RESULT 15
 A1824100/c 620 bp mRNA linear EST 21-DEC-1999
 LOCUS A1824100/c 620 bp mRNA linear EST 21-DEC-1999
 DEFINITION w146c01.x1 NCI-CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2405856 3'
 similar to TR:075667 D77116.4 ; contains PTR5.b2 PTR5
 repetitive element ; mRNA sequence.

ACCESSION A1824100
 VERSION A1824100.1 GI:5444771
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 620)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL, at:
 www-bio.lnl.gov/dbfp/image/image.html
 Insert length: 917 Std Error: 0.00
 Seq primer: -40UP from Glbco
 High quality sequence stop: 490.

FEATURES
 source Location/Qualifiers
 1..620
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2405856"
 /tissue="IMAGE:2405856"
 /clone_type="squamous cell carcinoma, poorly
 differentiated (4 pooled tumors, including primary and
 metastatic)"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /clone_lib="NCI-CGAP_Lu19"
 /note="Organ: lung; Vector: p773D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 pooled lung tumor tissue, and was then primed with a Not I
 - oligo(dT) primer. Double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 p773 vector. Library went through one round of
 normalization. Library constructed by Bento Soares and M.
 Fatima Bonaldo."

BASE COUNT 106 a 195 c 227 g 89 t 3 others

ORIGIN

Query Match 2.9%; Score 59; DB 9; Length 620;
 Best Local Similarity 100.0%; Pred. No. 1.9e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCGCTCTTACAGCAGCTGCTGCTCAAGAGGCGCTTTCAACCTGCAG 682
 |||||||
 DB 314 AGGCTGCGCGCTCTTACAGCAGCTGCTGCTCAAGAGGCGCTTTCAACCTGCAG 256

Search completed: August 15, 2003, 04:13:33
 Job time : 3888 secs

TT SMTT
EGADP
BLA
XMAK
USPSTU

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2003, 00:46:57 ; Search time 141 Seconds
(without alignments)
6395.360 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gattccatcgtctggtgta.....tggatcctcaaaaaaaaaa 2043

Scoring table: OLIGO_NUC

Gapop 60.0 , Capext 60.0

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 404290

Minimum DB seq length: 50

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database :
1: Issued Patents_NA:*
2: /cgn2_6/prodata/1/lna/5A_COMB.seq:*
3: /cgn2_6/prodata/1/lna/5B_COMB.seq:*
4: /cgn2_6/prodata/1/lna/5A_COMB.seq:*
5: /cgn2_6/prodata/1/lna/PCTUS_COMB.seq:*
6: /cgn2_6/prodata/1/lna/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score-distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1547	75.7	2032	3	US-09-045-284A-1
2	1547	75.7	2032	4	US-09-190-911-2
3	1153	56.4	2065	4	US-09-786-240-26
4	19	0.9	3792	2	US-08-992-334-1
5	19	0.9	3792	2	US-08-302-752-1
6	19	0.9	5234	2	US-08-992-334-2
7	19	0.9	5234	2	US-08-302-752-2
8	19	0.9	6722	2	US-08-992-334-3
9	19	0.9	6722	3	US-08-302-752-3
10	18	0.9	150	4	US-09-439-313-320
11	18	0.9	150	4	US-09-352-616A-320
12	18	0.9	150	4	US-09-232-149A-320
13	18	0.9	201	3	US-09-020-956-78
14	18	0.9	201	3	US-09-030-607-78
15	18	0.9	201	4	US-09-439-313-78
16	18	0.9	201	4	US-09-352-616A-78
17	18	0.9	201	4	US-09-232-149A-78
18	18	0.9	202	3	US-09-030-607-180
19	18	0.9	202	4	US-09-439-313-180
20	18	0.9	202	4	US-09-352-616A-180
21	18	0.9	202	4	US-09-232-149A-180
22	18	0.9	536	3	US-09-020-956-69
23	18	0.9	536	3	US-09-030-607-69
24	18	0.9	536	4	US-09-439-313-69
25	18	0.9	536	4	US-09-352-616A-69
26	18	0.9	536	4	US-09-232-149A-69
27	18	0.9	699	4	US-09-107-532A-1333

28	18	0.9	822	3	US-09-020-956-29	Sequence 29, App1
29	18	0.9	822	3	US-09-030-607-29	Sequence 29, App1
30	18	0.9	822	4	US-09-439-313-29	Sequence 29, App1
31	18	0.9	822	4	US-09-352-616A-29	Sequence 29, App1
32	18	0.9	822	4	US-09-232-149A-29	Sequence 29, App1
33	18	0.9	1197	4	US-09-257-179-26	Sequence 26, App1
34	18	0.9	1368	4	US-09-118-554-61	Sequence 61, App1
35	18	0.9	1368	4	US-09-118-554-61	Sequence 61, App1
36	18	0.9	1368	4	US-09-602-877A-61	Sequence 61, App1
37	18	0.9	1913	4	US-09-589-360B-40	Sequence 40, App1
38	18	0.9	1927	4	US-09-336-536-66	Sequence 66, App1
39	18	0.9	4351	4	US-09-634-238-25	Sequence 25, App1
40	17	0.8	141	1	US-08-036-555B-161	Sequence 161, App
41	17	0.8	141	1	US-08-469-569-161	Sequence 161, App
42	17	0.8	141	1	US-08-249-322A-161	Sequence 161, App
43	17	0.8	141	1	US-08-469-526A-161	Sequence 161, App
44	17	0.8	141	2	US-08-734-591A-161	Sequence 161, App
45	17	0.8	141	2	US-08-469-660-161	Sequence 161, App

ALIGNMENTS

RESULT 1									
US-09-045-284A-1									
Sequence 1, Application US/09045284A									
Patent No. 6265192									
GENERAL INFORMATION:									
APPLICANT: Bistup, Annette									
APPLICANT: Rosen, Steven D.									
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3									
FILE REFERENCE: 6510-107051									
CURRENT APPLICATION NUMBER: US/09/045,284A									
CURRENT FILING DATE: 1998-03-20									
NUMBER OF SEQ ID NOS: 9									
SOFTWARE: FastSeq for Windows Version 4.0									
SEQ ID NO 1									
LENGTH: 2032									
TYPE: DNA									
ORGANISM: Homo sapiens									
US-09-045-284A-1									
Query Match									
Best Local Similarity 99.9%; Pred. No. 0;									
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
75.7%; Score 1547; DB 3; Length 2032;									
QY	106	CAAGCTCTTCACCTTCAGCACATGCTACTGCTTAATAAATGAAGCTCCTGCTTTCT	165						
DB	160	CAAGCTCTTCACCTTCAGCACATGCTACTGCTTAATAAATGAAGCTCCTGCTTTCT	219						
QY	166	GTTTCCCAATGGCCATCTTGCTATCTTCACATGATGACGACCAACATCAGCTC	225						
DB	220	GTTTCCCAATGGCCATCTTGCTATCTTCACATGATGACGACCAACATCAGCTC	279						
QY	226	CTGTCTATGAAGCAGCCGAGCCATGACGCTGCTGCTTGTCTTCTGCGCCTC	285						
DB	280	CTGTCTATGAAGCAGCCGAGCCATGACGCTGCTGCTTGTCTTCTGCGCCTC	339						
QY	286	TGCTCTCTCTTGTGGGGAGCTTTTGGGACACCCAGATTTTCTACCTGATGA	345						
DB	340	TGCTCTCTCTTGTGGGGAGCTTTTGGGACACCCAGATTTTCTACCTGATGA	399						
QY	346	GCCGCTGCGACGCTGTGATGACCTTCAAGCAGACGCGCTGTGTCGACATGCC	405						
DB	400	GCCGCTGCGACGCTGTGATGACCTTCAAGCAGACGCGCTGTGTCGACATGCC	459						
QY	406	TGTGCGGATCTGATGAGGCGCTCTTGTGACATGAGCGCTTTGATGCGTCAT	465						
DB	460	TGTGCGGATCTGATGAGGCGCTCTTGTGACATGAGCGCTTTGATGCGTCAT	519						
QY	466	GGAACCTGTCCTCCGAGACAGTCCAGCTCTTTCAGTGGAGAACGCGGCGCTGTG	525						

```
Db 520 GGAACCTGTCCTCCCGAGACAGTCCAGCCTTTTTCAGTGGGAGAACAGCCGGGCGCTGTG 579
OY 526 TTCTGCACCTGCTGTGACATCATCCACAAAGATGAATCAATCCCCGGGCTACTGCAG 585
Db 580 TTCTGCACCTGCTGTGACATCATCCACAAAGATGAATCAATCCCCGGGCTACTGCAG 639
OY 586 GCTCTGTGACATCAACAGCCCTTTGAGGTGTGAGAAAGCCCTGCTCTACAGCA 645
Db 640 GCTCTGTGACATCAACAGCCCTTTGAGGTGTGAGAAAGCCCTGCTCTACAGCA 699
OY 646 GCTGTGCTCAAGAGGAGGTGCTTCTTCAACCTGAGTCCCTTCAACCCCTGCTGAAGA 705
Db 700 GCTGTGCTCAAGAGGAGGTGCTTCTTCAACCTGAGTCCCTTCAACCCCTGCTGAAGA 759
OY 706 CCCCTCCCTCAACCTGATATGCTGACACCTGCTCCGGAGCCCGGGCGGTGTCCGTT 765
Db 760 CCCCTCCCTCAACCTGATATGCTGACACCTGCTCCGGAGCCCGGGCGGTGTCCGTT 819
OY 766 CCGAGAACGACAAAGGAGATCTCATGATTGACAGTCCGATTTGTGATGGGCGAGCATGA 825
Db 820 CCGAGAACGACAAAGGAGATCTCATGATTGACAGTCCGATTTGTGATGGGCGAGCATGA 879
OY 826 GCAAAACCTCAAGAGGAGGACCAACCCACTATGATGACAGTCCATTCGCCAAAGCA 885
Db 880 GCAAAACCTCAAGAGGAGGACCAACCCACTATGATGACAGTCCATTCGCCAAAGCA 939
OY 886 GCTGAGATCTCAAGAGGACCAACCTGCTTCCCAAGGCGCTGACAGAAAGCTACCTGCT 945
Db 940 GCTGAGATCTCAAGAGGACCAACCTGCTTCCCAAGGCGCTGACAGAAAGCTACCTGCT 999
OY 946 TGTGGGCTATGAGAGCTGTGCTGAGCCCTGTGCGCCAGACTTCCCGAATGTATGANT 1005
Db 1000 TGTGGGCTATGAGAGCTGTGCTGAGCCCTGTGCGCCAGACTTCCCGAATGTATGANT 1059
OY 1006 CTTGGGATTTGATTTCTTCCCATCTTCTGACAGCTGGGAGCAATACATACCCGAGGCA 1065
Db 1060 CTTGGGATTTGATTTCTTCCCATCTTCTGACAGCTGGGAGCAATACATACCCGAGGCA 1119
OY 1066 GGGCATGGGTGACACAGCTTTCCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGC 1125
Db 1120 GGGCATGGGTGACACAGCTTTCCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGC 1179
OY 1126 TTGGGGCTGTGCTTGGCCCTATGAAAAGTTCTGAGTTCAGAAAAGCTGTGGCGATGC 1185
Db 1180 TTGGGGCTGTGCTTGGCCCTATGAAAAGTTCTGAGTTCAGAAAAGCTGTGGCGATGC 1239
OY 1186 CATGAATTTGCTGGGCTACCGCAGCTGATCTGAACAAAGAGAGGAGTGTGCT 1245
Db 1240 CATGAATTTGCTGGGCTACCGCAGCTGATCTGAACAAAGAGAGGAGTGTGCT 1299
OY 1246 GATCTTTCTGTCTACCTGAGCTGTCCCTGAGCAATCCACTAAGAGGTTGAGAGGCTT 1305
Db 1300 GATCTTTCTGTCTACCTGAGCTGTCCCTGAGCAATCCACTAAGAGGTTGAGAGGCTT 1359
OY 1306 TGTCTGCACCTGTGTGACGCTCAGTCACTTCTGATGCTGTGAGCCCTTGCTTACA 1365
Db 1360 TGTCTGCACCTGTGTGACGCTCAGTCACTTCTGATGCTGTGAGCCCTTGCTTACA 1419
OY 1366 TGTCTGAGCCTTAACATGCTGTGGTATCACACTGAGTGTGAGTGTCTCCACAG 1425
Db 1420 TGTCTGAGCCTTAACATGCTGTGGTATCACACTGAGTGTGAGTGTCTCCACAG 1479
OY 1426 TGTCTCAAGAGAGGACTTTGTGTCATGCTTTGTGTCTAGAAAACAGCTGGGAGACT 1485
Db 1480 TGTCTCAAGAGAGGACTTTGTGTCATGCTTTGTGTCTAGAAAACAGCTGGGAGACT 1539
OY 1486 TATGTGAGAGAGCATCCACAGTGAAGAGGATTTGCTTCTTCTTCTGATCT 1545
Db 1540 TATGTGAGAGAGCATCCACAGTGAAGAGGATTTGCTTCTTCTTCTGATCT 1599
OY 1546 TCTCTGTGGGAGACTTCAGAGACTTTGTGCTGTGAGAGGCTATTAAACAGAGACACT 1605
Db 1600 TCTCTGTGGGAGACTTCAGAGACTTTGTGCTGTGAGAGGCTATTAAACAGAGACACT 1659
```

```
OY 1606 ATCAGTGAATTGATTCATAAACCTCCCTGTCACATCTTGGCCAAATGGGAGATGATCT 1665
Db 1660 ATCAGTGAATTGATTCATAAACCTCCCTGTCACATCTTGGCCAAATGGGAGATGATCT 1719
OY 1666 TTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
Db 1720 TTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1757

RESULT 2
US-09-190-911-2
; Sequence 2, Application US/09190911
; Patent No. 636365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerlich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-2

Query Match 75.7%; Score 1547; DB 4; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

OY 106 CAAGTCTTCCACTTCAGACAAATGCTACTGCTTAAAAAATGAAAGCTCTGCTGTTCT 165
Db 160 CAAGTCTTCCACTTCAGACAAATGCTACTGCTTAAAAAATGAAAGCTCTGCTGTTCT 219
OY 166 GGTTCACGATGGCCATCTTGGCTATTTCCACATGTAAGGACACAAATCAGCTC 225
Db 220 GGTTCACGATGGCCATCTTGGCTATTTCCACATGTAAGGACACAAATCAGCTC 279
OY 226 CCTGTATGAAGGACACAGCCGAGGAGCATGACGTGCTGTCTCTCTGCGCTC 285
Db 280 CCTGTATGAAGGACACAGCCGAGGAGCATGACGTGCTGTCTCTCTGCGCTC 339
OY 286 TGGCTTTCTTTTGTGGGAGCTTTTGGGACAGCACCCAGATGTTTCTACTGATGA 345
Db 340 TGGCTTTCTTTTGTGGGAGCTTTTGGGAGCACCCAGATGTTTCTACTGATGA 399
OY 346 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGACACCGGCTGATGCTGACATGGC 405
Db 400 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGACACCGGCTGATGCTGACATGGC 459
OY 406 TGTGCGGATCTGATACGGGCGCTTCTTGTGCGACATGAGCGTCTTGTATGCTACAT 465
Db 460 TGTGCGGATCTGATACGGGCGCTTCTTGTGCGACATGAGCGTCTTGTATGCTACAT 519
OY 466 GGAACCTGTCCTCCCGAGAGCACTTCAAGCCTCTTCAAGTGGAGAAACAGCCGGGCTGTG 525
Db 520 GGAACCTGTCCTCCCGAGAGCACTTCAAGCCTCTTCAAGTGGAGAAACAGCCGGGCTGTG 579
OY 526 TTTGACACCTGCTGATCATCATCCACAAAGATGAATCAATCCCGGCTCACTGCAG 585
Db 580 TTTGACACCTGCTGATCATCATCCACAAAGATGAATCAATCCCGGCTCACTGCAG 639
OY 586 GCTCTGTGACATCAACAGCCCTTTGAGGTGTGAGAAAGCCCTGCTCTACAGCA 645
Db 640 GCTCTGTGACATCAACAGCCCTTTGAGGTGTGAGAAAGCCCTGCTCTACAGCA 699
```

646 CGTGTCTCAAGAGGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTAAGA 705
|||||
700 CGTGTCTCAAGAGGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTAAGA 759
706 CCCCTCCCTCAACCTGATATCGTGACCTGTGTCGGGAGCCCGGGCGGTGCCGTTC 765
|||||
760 CCCCTCCCTCAACCTGATATCGTGACCTGTGTCGGGAGCCCGGGCGGTGCCGTTC 819
766 CCGAGAACGACAAAGGAGATCATGATGACAGTCCGATGTGATGGGGGAGCATGA 825
|||||
820 CCGAGAACGACAAAGGAGATCATGATGACAGTCCGATGTGATGGGGGAGCATGA 879
826 GCAAACTCAAGAGGAGGACCAACCTTACTATGTATGAGTCACTGTGCAAGCA 885
|||||
880 GCAAACTCAAGAGGAGGACCAACCTTACTATGTATGAGTCACTGTGCAAGCA 939
886 GCTGGAATCTACAAAGACATCCAGTCTTGGCCCAAGGCGCTGAGGAGACCTTCT 945
|||||
940 GCTGGAATCTACAAAGACATCCAGTCTTGGCCCAAGGCGCTGAGGAGACCTTCT 999
946 TGTGCGCTATGAGAGCTGGCTGAGGCGCTGAGGCGAGCTCCGGAATGTATGATT 1005
|||||
1000 TGTGCGCTATGAGAGCTGGCTGAGGCGCTGAGGCGAGCTCCGGAATGTATGATT 1059
1006 CGTGGATTTGGAATTTTGGCCCATCTTCAAGACCTGGGTGATTAATCAACCCGAGCA 1065
|||||
1060 CGTGGATTTGGAATTTTGGCCCATCTTCAAGACCTGGGTGATTAATCAACCCGAGCA 1119
1066 GGGCATGGTGACACAGCTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGC 1125
|||||
1120 GGGCATGGTGACACAGCTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGC 1179
1126 TTTGGCGTGGCTTGTGGCCATGAAAGTTTCTGATTCAGAAAGCTGTGGCGATGC 1185
|||||
1180 TTTGGCGTGGCTTGTGGCCATGAAAGTTTCTGATTCAGAAAGCTGTGGCGATGC 1239
1186 CATGAATTTCTGTGCTGACCTGACCCAGCTGATCTGAACAGAGAAAGAACTGTTGCT 1245
1240 CATGAATTTCTGTGCTGACCTGACCCAGCTGATCTGAACAGAGAAAGAACTGTTGCT 1299
1246 GGAATCTTCTGTCTACCTGAGACTGTCCCTGAGCAATCCACTAAGAGGTTGAGAGGCTT 1305
1300 GGAATCTTCTGTCTACCTGAGACTGTCCCTGAGCAATCCACTAAGAGGTTGAGAGGCTT 1359
1306 TGGTGGCACCAGGAGTGTAGGCTGATCTTCTGTAAGTCTTCTGAGCCTTGCATCA 1365
1360 TGGTGGCACCAGGAGTGTAGGCTGATCTTCTGTAAGTCTTCTGAGCCTTGCATCA 1419
1366 TCTGTGAGCCTTAATCAATGATGTGTGGTATCAACAGTGAATGTGTGTGCACAG 1425
1420 TCTGTGAGCCTTAATCAATGATGTGTGGTATCAACAGTGAATGTGTGTGCACAG 1479
1426 TGTCTAAGCAGAGAGGACTTTTGTGTCTCATGCTGTGTCTAGAAACAGAGCTGGGAACT 1485
1480 TGTCTAAGCAGAGAGGACTTTTGTGTCTCATGCTGTGTCTAGAAACAGAGCTGGGAACT 1539
1486 TATGTGAGCAGACATCCCAAGGATGTAAGAGGATTTGCTTCTTCTTGTGATCT 1545
1540 TATGTGAGCAGACATCCCAAGGATGTAAGAGGATTTGCTTCTTCTTGTGATCT 1599
1546 TCTGTGCTGGGAGACTTTCAGAGACTTGTGGCCTGAGAGGCTTATTAAGCAGACAGT 1605
1600 TCTGTGCTGGGAGACTTTCAGAGACTTGTGGCCTGAGAGGCTTATTAAGCAGACAGT 1659
1606 ATCAGTGAATTTGATTCATAAACTCCCTGTCCACATCTTGGCCCAATGGGGAATGATCT 1665
1660 ATCAGTGAATTTGATTCATAAACTCCCTGTCCACATCTTGGCCCAATGGGGAATGATCT 1719
1666 TTTCAACAAAGAGTCAACAGATTTTCCACAGAGATGC 1703
1720 TTTCAACAAAGAGTCAACAGATTTTCCACAGAGATGC 1757

RESULT 3
US-09-786-240-26
; Sequence 26: Application US/09786240
; Patent No. 6558935
; GENERAL INFORMATION:
; APPLICANT: INCYTE PHARMACEUTICALS, INC.
; APPLICANT: TANG, Y. TOM
; APPLICANT: CORLEY, Neil C.
; APPLICANT: GUEGLER, Karl J.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LAL, Preeti
; APPLICANT: YOE, Henry
; APPLICANT: HILMAN, Jennifer L.
; APPLICANT: AZIMZAI, Valda
; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: P-0592 PCT
; CURRENT APPLICATION NUMBER: US/09/786,240
; PRIOR FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,6
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 26
; LENGTH: 2065
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: Incyte ID No. 6558935 2617407CB1
US-09-786-240-26

Query Match 56.4%; Score 1153; DB 4; Length 2065;
Best Local Similarity 99.6%; Pred. No. 0;
Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

106 CAAGTCTTCCACTTACAGACATGCTACTGCTTAAAAAATGAAAGTCTGCTGTTTCT 165
152 CAAGTCTTCCACTTACAGACATGCTACTGCTTAAAAAATGAAAGTCTGCTGTTTCT 211
166 GGTTCACAGATGGCCATCTTGGCTCTATTCTTCACATGTACAGCCCAACATCAGCTC 225
212 GGTTCACAGATGGCCATCTTGGCTCTATTCTTCACATGTACAGCCCAACATCAGCTC 271
226 CCTGTATGAAGGACAGCCGAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 285
272 CCTGTATGAAGGACAGCCGAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 331
286 TGGCTCTCTTTTGGGGAGCTTTTGGGGAGCAGACCCAGATGTTTCTACCTGATGA 345
332 TGGCTCTCTTTTGGGGAGCTTTTGGGGAGCAGACCCAGATGTTTCTACCTGATGA 391
346 GCCCGCTGGCAGCTGTGATGACCTTCAACAGACAGCCCTGTGATGTGCATAGGC 405
392 GCCCGCTGGCAGCTGTGATGACCTTCAACAGACAGCCCTGTGATGTGCATAGGC 451
406 TGTGCGGATCTGATGATGAGGCGCTTCTTGTGGACATGAGCGTGTGATGCTCAT 465
452 TGTGCGGATCTGATGATGAGGCGCTTCTTGTGGACATGAGCGTGTGATGCTCAT 511
466 GGAACCTGTGCTCCGGAGACAGTCCAGCTTTCAGTGGGAGAAACAGCCGGGCGCTGTG 525
512 GGAACCTGTGCTCCGGAGACAGTCCAGCTTTCAGTGGGAGAAACAGCCGGGCGCTGTG 571
526 TTTGTGACCTCCCTGTGACATCATCCCAAGATGAATCATCCCGGGGCTCACTGAG 585
572 TTTGTGACCTCCCTGTGACATCATCCCAAGATG -AATCATCCCGGGGCTCACTGAG 630
586 GCTCCTGTGAGTGAAGAGCCCTTGTG -AGTGTGAGGAAGAGCGCTGCGCTCTACAGCC 644
631 GCTCCTGTGAGTGAAGAGCCCTTGTGAGTGTGAGGAAGAGCGCTGCGCTCTACAGCC 690
645 AGTGTGCTTCAAGAGGAGTGGCTTCTTCAACCTGACAGTCCCTCTACCCGCTGTGAAG 704
|||||

```
Db 691 AGGTGTCCTCAAGAGAGTGGCTTCTTCAACCTGACGTCCTTACCCGCTGCTGAAG 750
OY 705 ACCCTCCCTCAACCTGCATATGTCACCTGCTGCGGAGCCCGGCGGCTTCGTT 764
Db 751 ACCCTCCCTCAACCTGCATATGTCACCTGCTGCGGAGCCCGGCGGCTTCGTT 810
OY 765 CCCGAGAACGCAAAAGGAGATCTCANGATTGACAGTGCATTTGATGGGCGACATG 824
Db 811 CCCGAGAACGCAAAAGGAGATCTCANGATTGACAGTGCATTTGATGGGCGACATG 870
OY 825 ACCAAACTCAAGAGAGAGCAACCCCTACTATGATGATGAGAGCATCTGCCAAGCC 884
Db 871 ACCAAACTCAAGAGAGAGCAACCCCTACTATGATGATGAGAGCATCTGCCAAGCC 930
OY 885 ACCTGAGATCTACAAGACCATTCAGTCCCTTGCACAGGCGCTGCAGAACCTTACTGC 944
Db 931 ACCTGAGATCTACAAGACCATTCAGTCCCTTGCACAGGCGCTGCAGAACCTTACTGC 990
OY 945 TTGTCGCTATAGAGACCTGGCTCGAGCCCTGTCGCGCCAGACCTTCCGAAATGATGAT 1004
Db 991 TTGTCGCTATAGAGACCTGGCTCGAGCCCTGTCGCGCCAGACCTTCCGAAATGATGAT 1050
OY 1005 TCGTGGGATTTGGAATTTCTTGGCCCACTTCAGACCTGGGTGATATACATCACCCGAGCA 1064
Db 1051 TCGTGGGATTTGGAATTTCTTGGCCCACTTCAGACCTGGGTGATATACATCACCCGAGCA 1110
OY 1065 AGGCAATGGGTGACCAAGCTTTCACACAATGCCAGGAGTCCCTTAATGTCTCCAG 1124
Db 1111 AGGCAATGGGTGACCAAGCTTTCACACAATGCCAGGAGTCCCTTAATGTCTCCAG 1170
OY 1125 CTTGGCGCTGCTTTCCTTGAAGAGTTCTTCGATTTGAGAAAGCTGTGGGAGT 1184
Db 1171 CTTGGCGCTGCTTTCCTTGAAGAGTTCTTCGATTTGAGAAAGCTGTGGGAGT 1230
OY 1185 CCATGAATTTGCTGGGCTACCCGCAAGTCAGATCTGACAAAGAGAGAAACCTTGTGC 1244
Db 1231 CCATGAATTTGCTGGGCTACCCGCAAGTCAGATCTGACAAAGAGAGAAACCTTGTGC 1290
OY 1245 TGGATCTTCTGTCACCTGAGCTGTCCGAGCAATTCACATGAGAGGCTTGAGAGGCT 1304
Db 1291 TGGATCTTCTGTCACCTGAGCTGTCCGAGCAATTCACATGAGAGGCTTGAGAGGCT 1350
OY 1305 TTGCTGCCACCTGATGTCAGCTCACTTCTCTGATGCTTCTGAGCCCTTGGCTAC 1364
Db 1351 TTGCTGCCACCTGATGTCAGCTCACTTCTCTGATGCTTCTGAGCCCTTGGCTAC 1410
OY 1365 ATCTCTGAGCCTTAATACATGTCGTGGGATTCACACTGAGTGTGATGTCACAC 1424
Db 1411 ATCTCTGAGCCTTAATACATGTCGTGGGATTCACACTGAGTGTGATGTCACAC 1470
OY 1425 GTGCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTAGAAAACAGACTGGGAGCC 1484
Db 1471 GTGCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTAGAAAACAGACTGGGAGCC 1530
OY 1485 TTATGTAGACACACATCCACCAAGAGAGGATTTGCTTCTTTCTTTCTGATC 1544
Db 1531 TTATGTAGACACACATCCACCAAGAGAGGATTTGCTTCTTTCTTTCTGATC 1590
OY 1545 TTCTGTCTGGGAGAGCTTCAAGAGACTTTGTGGCTGAGAGGCTATTAAAGCAGACAG 1604
Db 1591 TTCTGTCTGGGAGAGCTTCAAGAGACTTTGTGGCTGAGAGGCTATTAAAGCAGACAG 1650
OY 1605 TTATAGTGAATTTGATCCATAAAGCTCCCTGTCCACATTTGCCCAATGGGAGATGATC 1664
Db 1651 TTATAGTGAATTTGATCCATAAAGCTCCCTGTCCACATTTGCCCAATGGGAGATGATC 1710
OY 1665 TTTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
Db 1711 TTTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1749
```

RESULT 4
US-08-992-334-1

```
Sequence 1, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Priout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: YES
IMMEDIATE SOURCE:
CLONE: pg+host4
US-08-992-334-1
Query Match 0.9%; Score 19; DB 2; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1821 TCTCTGACAAAGAGCAAG 1839
Db 1818 TCTCTGACAAAGAGCAAG 1836
RESULT 5
US-08-302-752-1
Sequence 1, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
```

```

: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25 (EFO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/302,752
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 9203034
: FILING DATE: 13-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO FR/93/00248
: FILING DATE: 12-MAR-1993
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 3792 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-302-752-1

Query Match          0.9%; Score 19; DB 3; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1821 TCCTGCAAAAGAGCAAG 1839
Db      1818 TCCTGCAAAAGAGCAAG 1836
|||||

RESULT 6
: Sequence 2, Application US/08992334
: Patent No. 5919678
: GENERAL INFORMATION:
: APPLICANT: Gruss, Alexandra
: APPLICANT: Maguin, Emmanuelle
: TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
: TITLE OF INVENTION: PLASMID
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Christie Parker & Hale, LLP
: STREET: 350 West Colorado Boulevard, Suite 500
: CITY: Pasadena
: STATE: California
: COUNTRY: United States
: ZIP: 91105
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/992,334
: FILING DATE: 17-DEC-1997
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/302,752
: FILING DATE: 24-DEC-1994
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: PCT/FR93/00248
: FILING DATE: 12-MAR-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR FR92/03034
: FILING DATE: 13-MAR-1992
: ATTORNEY/AGENT INFORMATION:
: NAME: Proulx, D. Bruce
: REGISTRATION NUMBER: 20958
: REFERENCE/DOCKET NUMBER: C93.31779
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (626) 795-9900
: TELEFAX: (626) 577-8800
: INFORMATION FOR SEQ ID NO: 2:
```

```

: SEQUENCE CHARACTERISTICS:
: LENGTH: 5234 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: both
: TOPOLOGY: circular
: MOLECULE TYPE: DNA (genomic)
: US-08-992-334-2

Query Match          0.9%; Score 19; DB 2; Length 5234;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1821 TCCTGCAAAAGAGCAAG 1839
Db      3260 TCCTGCAAAAGAGCAAG 3278
|||||

RESULT 7
: Sequence 2, Application US/08302752
: Patent No. 6025190
: GENERAL INFORMATION:
: APPLICANT:
: TITLE OF INVENTION: THERMOSENSIBLE PLASMID
: NUMBER OF SEQUENCES: 3
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25 (EFO)
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/302,752
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: FR 9203034
: FILING DATE: 13-MAR-1992
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: WO FR/93/00248
: FILING DATE: 12-MAR-1993
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 5234 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: US-08-302-752-2

Query Match          0.9%; Score 19; DB 3; Length 5234;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1821 TCCTGCAAAAGAGCAAG 1839
Db      3260 TCCTGCAAAAGAGCAAG 3278
|||||

RESULT 8
: Sequence 3, Application US/08992334
: Patent No. 5919678
: GENERAL INFORMATION:
: APPLICANT: Gruss, Alexandra
: APPLICANT: Maguin, Emmanuelle
: TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
: TITLE OF INVENTION: PLASMID
: NUMBER OF SEQUENCES: 3
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Christie Parker & Hale, LLP
: STREET: 350 West Colorado Boulevard, Suite 500
: CITY: Pasadena
: STATE: California
```

```

; COUNTRY: United States
; ZIP: 91105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; .APPLICATION NUMBER: US/08/992,334
; FILING DATE: 17-DEC-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/302,752
; FILING DATE: 24-DEC-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/FR93/00248
; FILING DATE: 12-MAR-1993
; APPLICATION DATA:
; APPLICATION NUMBER: FR 992/03034
; FILING DATE: 13-MAR-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Prout, D. Bruce
; REGISTRATION NUMBER: 20958
; REFERENCE/DOCKET NUMBER: C93:31779
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (626) 795-9900
; TELEFAX: (626) 577-8800
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; US-08-992-334-3

Query Match
Best Local Similarity 100.0%; Score 19; DB 2; Length 6722;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCTCTGACAAAGAGCAG 1839
DB 4748 TCTCTGACAAAGAGCAG 4766

RESULT 9
; Sequence 3, Application US/08302752
; Patent No. 6025190
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: THERMOSENSIBLE PLASMID
; NUMBER OF SEQUENCES: 3
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/302,752
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: FR 9203034
; FILING DATE: 13-MAR-1992
; APPLICATION DATA:
; APPLICATION NUMBER: WO FR/93/00248
; FILING DATE: 12-MAR-1993
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 6722 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
```

```

; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-302-752-3

Query Match
Best Local Similarity 100.0%; Score 19; DB 3; Length 6722;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCTCTGACAAAGAGCAG 1839
DB 4748 TCTCTGACAAAGAGCAG 4766

RESULT 10
; US-09-439-313-320
; Sequence 320, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yuqul
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Ranger, Gary
; APPLICANT: Retter, Mark
; APPLICANT: Solk, John
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 320
; LENGTH: 150
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-439-313-320

Query Match
Best Local Similarity 100.0%; Score 18; DB 4; Length 150;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCATTGTTGG 18
DB 31 GAATTCATTGTTGG 48

RESULT 11
; US-09-352-616A-320
; Sequence 320, Application US/09352616A
; Patent No. 6395278
; GENERAL INFORMATION:
; APPLICANT: Dillon, Davin C.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang, Yuqul
; APPLICANT: Xu, Jiangchun
; APPLICANT: Mitcham, Jennifer Lynn
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
; FILE REFERENCE: 210121.427C8
; CURRENT APPLICATION NUMBER: US/09/352,616A
; NUMBER OF SEQ ID NOS: 472
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 320
; LENGTH: 150
; TYPE: DNA
; ORGANISM: Homo sapien
; US-09-352-616A-320
```

Query Match 0.9%; Score 18; DB 4; Length 150;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCATTGTGTGGG 18
|||||
DB 31 GAATTCATTGTGTGGG 48

RESULT 12

US-09-232-149A-320
; Sequence 320, Application US/09232149A
; Patent No. 6465611
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
; FILE REFERENCE: 210121.427C6
; CURRENT APPLICATION NUMBER: US/09/232.149A
; CURRENT FILING DATE: 1999-01-15
; NUMBER OF SEQ ID NOS: 338
; SOFTWARE: FASTSEQ for Windows Version 3.0
; SEQ ID NO 320
; LENGTH: 150
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-232-149A-320

Query Match 0.9%; Score 18; DB 4; Length 150;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCATTGTGTGGG 18
|||||
DB 31 GAATTCATTGTGTGGG 48

RESULT 13

US-09-020-956-78
; Sequence 78, Application US/09020956
; Patent No. 6261562
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS FO
; NUMBER OF SEQUENCES: 178
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/020.956
; FILING DATE: 09-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 78:

SEQUENCE CHARACTERISTICS:

LENGTH: 201 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-09-020-956-78

Query Match 0.9%; Score 18; DB 3; Length 201;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCATTGTGTGGG 18
|||||
DB 18 GAATTCATTGTGTGGG 35

RESULT 14

US-09-030-607-78
; Sequence 78, Application US/09030607
; Patent No. 6262245
; GENERAL INFORMATION:
; APPLICANT: Xu, Jiangchun
; APPLICANT: Dillon, David C.
; TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
; NUMBER OF SEQUENCES: 224
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: SEED AND BERRY LLP
; STREET: 6300 Columbia Center, 701 Fifth Avenue
; CITY: Seattle
; STATE: WA
; COUNTRY: USA
; ZIP: 98104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/030.607
; FILING DATE: 25-FEB-1998
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Makl, David J.
; REGISTRATION NUMBER: 31,392
; REFERENCE/DOCKET NUMBER: 210121.427C3
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (206) 622-4900
; TELEFAX: (206) 682-6031
; INFORMATION FOR SEQ ID NO: 78:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 201 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; ORIGINAL SOURCE:
; ORGANISM: Homo sapiens
US-09-030-607-78

Query Match 0.9%; Score 18; DB 3; Length 201;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCATTGTGTGGG 18
|||||
DB 18 GAATTCATTGTGTGGG 35

RESULT 15
US-09-439-313-78

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2003, 00:56:07 ; Search time 1505 Seconds
(without alignments)
3037.948 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaaatccatgtgtgtgta.....tgggacccaataaaaaa 2043

Scoring table:

OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Searched: 1504479 seqs, 1118970152 residues

Word size : 0

Total number of hits satisfying chosen parameters: 2057348

Minimum DB seq length: 50

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database : Published_Applications_NA.*

1: /cgn2_6/ptodata/2/pubpna/US07_PUBCOMB.seq.*
2: /cgn2_6/ptodata/2/pubpna/PCT_NEW_PUB.seq.*
3: /cgn2_6/ptodata/2/pubpna/US06_NEW_PUB.seq.*
4: /cgn2_6/ptodata/2/pubpna/US05_PUBCOMB.seq.*
5: /cgn2_6/ptodata/2/pubpna/US07_NEW_PUB.seq.*
6: /cgn2_6/ptodata/2/pubpna/PCTUS_PUBCOMB.seq.*
7: /cgn2_6/ptodata/2/pubpna/US08_NEW_PUB.seq.*
8: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
9: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
10: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
11: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
12: /cgn2_6/ptodata/2/pubpna/US09_PUBCOMB.seq.*
13: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
14: /cgn2_6/ptodata/2/pubpna/US10_PUBCOMB.seq.*
15: /cgn2_6/ptodata/2/pubpna/US10_NEW_PUB.seq.*
16: /cgn2_6/ptodata/2/pubpna/US60_NEW_PUB.seq.*
17: /cgn2_6/ptodata/2/pubpna/US60_PUBCOMB.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1547	75.7	2032	9 US-09-816-825-1
2	1547	75.7	2032	13 US-10-007-262-2
3	403	19.7	505	10 US-09-998-598-2595
4	403	19.7	517	10 US-09-919-580-194
5	114	5.6	389	10 US-09-919-580-264
6	59	2.9	2544	9 US-09-927-602-1
7	59	2.9	48436	9 US-09-927-602-38
8	22	1.1	167	10 US-09-867-701-9458
9	22	1.1	171	10 US-09-867-701-9458
10	22	1.1	187	10 US-09-867-701-9067
11	22	1.1	191	10 US-09-867-701-9185
12	20	1.0	398	10 US-09-960-352-11099
13	20	1.0	570	10 US-10-027-632-140660
14	20	1.0	11102	14 US-10-205-823-334
15	20	1.0	180557	13 US-10-003-806-6
16	20	1.0	180557	13 US-10-003-806-9

C 17	19	0.9	341	10	US-09-867-701-8841	Sequence 8841, Ap
C 18	19	0.9	341	10	US-09-867-701-8888	Sequence 8888, Ap
C 19	19	0.9	349	9	US-09-759-143-649	Sequence 649, Ap
C 20	19	0.9	349	9	US-09-780-669-649	Sequence 649, Ap
C 21	19	0.9	349	9	US-09-822-827-649	Sequence 649, Ap
C 22	19	0.9	349	9	US-09-895-793-649	Sequence 649, Ap
C 23	19	0.9	349	10	US-09-895-814-649	Sequence 649, Ap
C 24	19	0.9	349	13	US-10-012-896-649	Sequence 649, Ap
C 25	19	0.9	405	10	US-09-867-701-9634	Sequence 9634, Ap
C 26	19	0.9	544	13	US-10-027-632-87347	Sequence 87347, A
C 27	19	0.9	766	13	US-10-027-632-118925	Sequence 118925,
C 28	19	0.9	2874	13	US-10-027-632-111785	Sequence 111785,
C 29	19	0.9	3273	12	US-09-971-392-26	Sequence 26, Ap
C 30	19	0.9	3274	13	US-10-002-600-52	Sequence 52, Ap
C 31	19	0.9	4234	19	US-09-815-242-9944	Sequence 9944, Ap
C 32	19	0.9	4537	13	US-10-108-605-322	Sequence 322, Ap
C 33	19	0.9	6346	12	US-10-017-161-2201	Sequence 2201, Ap
C 34	19	0.9	8742	11	US-09-908-711-147	Sequence 147, Ap
C 35	19	0.9	8742	11	US-09-764-891-6217	Sequence 6217, Ap
C 36	19	0.9	10929	14	US-10-032-393-7	Sequence 7, Ap
C 37	19	0.9	62944	10	US-09-954-456-2257	Sequence 2257, Ap
C 38	19	0.9	75899	9	US-09-854-883-243	Sequence 243, Ap
C 39	19	0.9	684973	10	US-09-263-959-1	Sequence 1, Ap
C 40	18	0.9	129	9	US-09-759-143-765	Sequence 765, Ap
C 41	18	0.9	129	9	US-09-780-669-765	Sequence 765, Ap
C 42	18	0.9	129	9	US-09-822-827-765	Sequence 765, Ap
C 43	18	0.9	129	10	US-09-895-793-765	Sequence 765, Ap
C 44	18	0.9	129	10	US-09-895-814-765	Sequence 765, Ap
C 45	18	0.9	129	13	US-10-012-896-765	Sequence 765, Ap

ALIGNMENTS

RESULT 1
US-09-816-825-1
Sequence 1, Application US/09816825
Patent No. US20010051370A1
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107CON
CURRENT APPLICATION NUMBER: US/09/816, 825
CURRENT FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/045, 284
PRIOR FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2032
TYPE: DNA
ORGANISM: Homo sapiens
US-09-816-825-1

Query Match	75.7%	Score 1547	DB 9	Length 2032
Best Local Similarity	99.9%	Pred. No. 0		
Matches 1597	Conservative	0	Mismatches	1
			Indels	0
			Gaps	0
QY 106	CAAGGCTTCACATTCAGCACAATGCTACGCTAAATAAATGAGTCTGCTTTC	165		
DB 160	CAAGGCTTCACATTCAGCACAATGCTACGCTAAATAAATGAGTCTGCTTTC	219		
QY 166	GGTTTCCAGATGGCATCTTGGCTATCTTCACATGTACAGCCACAAATCAGCTC	225		
DB 220	GGTTTCCAGATGGCATCTTGGCTATCTTCACATGTACAGCCACAAATCAGCTC	279		
QY 226	CCGTGTATGACGACAGCCGACGATGACGCTGCTGCTCTTCTGCGGCTC	285		
DB 280	CCGTGTATGACGACAGCCGACGATGACGCTGCTGCTCTTCTGCGGCTC	339		
QY 286	TGGCTCTTCTTCTTCTGTTGGGAGAGCTTTTGGGACGACACCCAGATCTTTCTACTGATGGA	345		

Db	340	TGGCTCTCTTTTGGGGGAGCTTTTGGGAGAGACCAGATGTTTCTACGTGATGA	399
Oy	346	GGCCGCGTGGAGCGTGGGATGACCTTCAACAGACACCGCCGTGGATGCTGCACATGGC	405
Db	400	GGCCGCGCTGGAGCGTGGATGACCTTCAACAGACACCGCCGTGGATGCTGCACATGGC	459
Oy	406	TGTGCGGGATCTGATACGGGCGCTTCTTGTGGACATGAGCGCTTGTGATGCTACAT	465
Db	460	TGTGCGGGATCTGATACGGGCGCTTCTTGTGGACATGAGCGCTTGTGATGCTACAT	519
Oy	466	GGAACCTGTGTCCCGGAGACAGTCACGCTTCTTCAGTGGAGAAACCGCGCGCTGTG	525
Db	520	GGAACCTGTGTCCCGGAGACAGTCACGCTTCTTCAGTGGAGAAACACCGCGCGCTGTG	579
Oy	526	TTTGCACCTGCTGTGCATCATCCCAAGATGAATCATCCCGCGGGCTCACTGGAG	585
Db	580	TTTGCACCTGCTGTGCATCATCCCAAGATGAATCATCCCGCGGGCTCACTGGAG	639
Oy	586	GCTCCCTGCACTCAACAGCCCTTGTAGGTGTGGAGAGGCGTGGCGCTCTACAGCA	645
Db	640	GCTCCCTGCACTCAACAGCCCTTGTAGGTGTGGAGAGGCGTGGCGCTCTACAGCA	699
Oy	646	CGTGGTGTCAGAGGAGTGGCGCTTCTTCAACTGCAGTCCCTTCAACCGGTGCGTAAGA	705
Db	700	CGTGGTGTCAGAGGAGTGGCGCTTCTTCAACTGCAGTCCCTTCAACCGGTGCGTAAGA	759
Oy	706	CCCTCCCTCACTGCATATGTGCACCTGTCCGGAGCCCGGGCGCGTTCCTGCTTC	765
Db	760	CCCTCCCTCACTGCATATGTGCACCTGTCCGGAGCCCGGGCGCGTTCCTGCTTC	819
Oy	766	CCGGAACGCGCAAGGGAGATCTCATGTGACAGTCCGACATTTGATGGGGGCGACATGA	825
Db	820	CCGGAACGCGCAAGGGAGATCTCATGTGACAGTCCGACATTTGATGGGGGCGACATGA	879
Oy	826	GCAAAAACCTCAAGAGAGAGACCAACCTACTATGTGATGACAGTCACTGCCAAAGCA	885
Db	880	GCAAAAACCTCAAGAGAGAGACCAACCTACTATGTGATGACAGTCACTGCCAAAGCA	939
Oy	886	GCTGGAGATCTACAAGAACATCCAGTCTCTTGGCCCAAGGCCCTGCAGAGAACCTACTGCT	945
Db	940	GCTGGAGATCTACAAGAACATCCAGTCTCTTGGCCCAAGGCCCTGCAGAGAACCTACTGCT	999
Oy	946	TGTGCGCTATAGACACTGGCTGCAGGCCCTGTGGGCCAGACTTCCCAATGTATGAATTT	1005
Db	1000	TGTGCGCTATAGACACTGGCTGCAGGCCCTGTGGGCCAGACTTCCCAATGTATGAATTT	1059
Oy	1006	CGTGGGATTGAATTTCTTGCCCAATCTTTCAGACCTGGGTGCATTAACATCACCCGAGCAA	1065
Db	1060	CGTGGGATTGAATTTCTTGCCCAATCTTTCAGACCTGGGTGCATTAACATCACCCGAGCAA	1119
Oy	1066	GGGATGGGATACACAGCTTTCACACAAATGGCAGAGGATCCCTTAATGTCTCCAGGC	1125
Db	1120	GGGATGGGATACACAGCTTTCACACAAATGGCAGAGGATCCCTTAATGTCTCCAGGC	1179
Oy	1126	TTGGCGCTGCTTGTGCCCTATGAAAAGGTTTTCGACTTCAGAAAGCCTGTGGCGATGC	1185
Db	1180	TTGGCGCTGCTTGTGCCCTATGAAAAGGTTTTCGACTTCAGAAAGCCTGTGGCGATGC	1239
Oy	1186	CATGAATTTGCTGGGCTAACCGCCACAGTCAAGTCTGAACAAACAGAGAAACCTGTGTCT	1245
Db	1240	CATGAATTTGCTGGGCTAACCGCCACAGTCAAGTCTGAACAAACAGAGAAACCTGTGTCT	1299
Oy	1246	GGATCTTCTGTCTACCTGGAGTGTGCCCTGAGACCAATCCACTAAGAGGTTGAGAAGGCTT	1305
Db	1300	GGATCTTCTGTCTACCTGGAGTGTGCCCTGAGACCAATCCACTAAGAGGTTGAGAAGGCTT	1359
Oy	1306	TGCTGCCACCTGTGTGACGCTCAAGTCACTTCTCTGAATGCTTCTGAGCCTTGCTCAACA	1365
Db	1360	TGCTGCCACCTGTGTGACGCTCAAGTCACTTCTCTGAATGCTTCTGAGCCTTGCTCAACA	1419
Oy	1366	TTCTGTGAGCCTTAATCAATGTCTGTGGGTATCACACTGAGTGTGAGTGTGTGTCAACAG	1425

Db	1420	TCCTGAGACCCTTA	CTACATCTCTGTGGGTATACACAGATGTGAGTGTGTCCACAG	1479
Qy	1426	TGCTCAGACAGAA	GAGCTTTGTGTCCATGCTTGTGTCTAGAAAAAGACACTGGGAACT	1485
Db	1480	TGCTTAAGCACA	AAAGACTTTTGTGTCTCATGTGTGTCTAGAAAAAGACACTGGGAACT	1539
Qy	1486	TATGTGAGCAGAC	ATATCCACAGTGAACAGGGTATTTGCTCTTCTTTCTTGATCT	1545
Db	1540	TATGTGAGCAGAC	ATATCCACAGTGAACAGGGTATTTGCTCTTCTTTCTTGATCT	1599
Qy	1546	TCCTGTCTGGGCA	CACTTCAGACATTTGTGTGCTTGGAGGCTTTTAAGCACGACAGT	1605
Db	1600	TCCTGTCTGGGCA	CACTTCAGACATTTGTGTGCTTGGAGGCTTTTAAGCACGACAGT	1659
Qy	1606	ATCATGGAAATTA	ATGATCCATTAACCTCCCTGTCCACATCTTCCCAATGGGAAATGATCT	1665
Db	1660	ATCATGGAAATTA	ATGATCCATTAACCTCCCTGTCCACATCTTCCCAATGGGAAATGATCT	1719
Qy	1666	TTCAACCAAGAGC	TCACCGACTTTTCCACAGAGATGC	1703
Db	1720	TTCAACCAAGAGC	TCACCGACTTTTCCACAGAGATGC	1757
RESULT 2				
US-10-007-262-2				
Sequence 2, Application US/10007262				
Publication No. US20020164748a1				
GENERAL INFORMATION:				
APPLICANT: Bistrup, Annette				
APPLICANT: Rosen, Steven D.				
APPLICANT: Tangemann, Kirsten				
APPLICANT: Hammerich, Stefan				
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3				
FILE REFERENCE: 6510-107CIP				
CURRENT APPLICATION NUMBER: US/10/007,262				
PRIORITY FILING DATE: 2001-11-08				
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911				
PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12				
NUMBER OF SEQ ID NOS: 8				
SOFTWARE: FastSeq for Windows Version 3.0				
SEQ ID NO 2				
LENGTH: 2032				
TYPE: DNA				
ORGANISM: H. sapiens				
US-10-007-262-2				
Query Match				
Best Local Similarity 99.9%; Score 1547; DB 13; Length 2032;				
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;				
Qy	106	CAAGGTCTTCAC	TTGAGCACAAATGCTATGCTCTAAAAAATGAAAGCTCCGTGTTTCT	165
Db	160	CAAGGTCTTCAC	TTGAGCACAAATGCTATGCTCTAAAAAATGAAAGCTCCGTGTTTCT	219
Qy	166	GTTTCCAGATG	GCATCTTGCTCTATTTCCACATGTAGACCAACAACATCAGCTC	225
Db	220	GTTTCCAGATG	GCATCTTGCTCTATTTCCACATGTAGACCAACAACATCAGCTC	279
Qy	226	CCCTCTATATGA	AGACACAGCCCGAGCGCATCAGTGTGTTCTTCTTCCCTGGCCTC	285
Db	280	CCCTCTATATGA	AGACACAGCCCGAGCGCATCAGTGTGTTCTTCTTCCCTGGCCTC	339
Qy	286	TGCGTCTTCTTT	TGTGGGGCAGCTTTTGTGGGACACACCAAGATTTTCTACATGATGA	345
Db	340	TGCGTCTTCTTT	TGTGGGGCAGCTTTTGTGGGACACACCAAGATTTTCTACATGATGA	399
Qy	346	GCCCGCTGCGA	CTGTGATGACCTTCAACAGCACAGCCGCTGATGCTGCACATGGC	405
Db	400	GCCCGCTGCGA	CTGTGATGACCTTCAACAGCACAGCCGCTGATGCTGCACATGGC	459
Qy	406	TGTGCGGAGAT	GTATAGGGCGGCTCTTGTGGAGACATGAGCGCTTTGATGGCTACAT	465
Db	460	TGTGCGGAGAT	GTATAGGGCGGCTCTTGTGGAGACATGAGCGCTTTGATGGCTACAT	519

QY 466 GGAACCTGGTCCCGGAGACAGTCCAGCCTCTTTCAGTGGAGAAACACCCGGGCCCTGTG 525
DB 520 GGAACCTGGTCCCGGAGACAGTCCAGCCTCTTTCAGTGGAGAAACACCCGGGCCCTGTG 579
QY 526 TTCGTGACCTGCTGTGATCATATCCCAAGATGAATCATCCCGGGCTCACTGCAG 585
DB 580 TTCGTGACCTGCTGTGATCATATCCCAAGATGAATCATCCCGGGCTCACTGCAG 639
QY 586 GCTCTGTGAGTCAACAGCCCTTTGAGTGGTGAAGGCTGCGCTCTCAAGCA 645
DB 640 GCTCTGTGAGTCAACAGCCCTTTGAGTGGTGAAGGCTGCGCTCTCAAGCA 699
QY 646 CGTGTGCTCAAGAGAGTGTGCTTCTTCACTGAGTCCCTCAACCTGCTGCTGCAAGA 705
DB 700 CGTGTGCTCAAGAGAGTGTGCTTCTTCACTGAGTCCCTCAACCTGCTGCTGCAAGA 759
QY 706 CCCCTCCCTCAACCTGATATGCTGACCTGTGCTCCGGAGCCCGGGCTGTCCGTTC 765
DB 760 CCCCTCCCTCAACCTGATATGCTGACCTGTGCTCCGGAGCCCGGGCTGTCCGTTC 819
QY 766 CCGAAGAACGCAAGAGAGAGATCAGATGAGATGAGATGAGATGAGATGAGATGAGAT 825
DB 820 CCGAAGAACGCAAGAGAGAGATCAGATGAGATGAGATGAGATGAGATGAGATGAGAT 879
QY 826 GCAAAACTCAAGAGAGAGAGACCAACCTACTATGATGAGATGAGATGAGATGAGAT 885
DB 880 GCAAAACTCAAGAGAGAGAGACCAACCTACTATGATGAGATGAGATGAGATGAGAT 939
QY 886 GCTGAGATCTCAAGAGACCAACCTGCTTCCGAGGCTCCGAGAGACCTCACTGCT 945
DB 940 GCTGAGATCTCAAGAGACCAACCTGCTTCCGAGGCTCCGAGAGACCTCACTGCT 999
QY 946 TGTGGGCTATGAGAGACCTGCTGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGCT 1005
DB 1000 TGTGGGCTATGAGAGACCTGCTGAGGCTGCTGAGGCTGCTGAGGCTGCTGAGGCT 1059
QY 1006 CGTGGATTTGGAATCTTGGCCCATCTTCAAGACCTGCTGATTAATCACTCCAGGCA 1065
DB 1060 CGTGGATTTGGAATCTTGGCCCATCTTCAAGACCTGCTGATTAATCACTCCAGGCA 1119
QY 1066 GGGCATGGGTACCAACGCTTTCCACAAATGCGGAGATGCCCTTAATGCTCCAGGC 1125
DB 1120 GGGCATGGGTACCAACGCTTTCCACAAATGCGGAGATGCCCTTAATGCTCCAGGC 1179
QY 1126 TTGGGCTGCTTGTGCTTGAAGAGTTCGATGAGAGAGCTGCTGAGGCTGCT 1185
DB 1180 TTGGGCTGCTTGTGCTTGAAGAGTTCGATGAGAGAGCTGCTGAGGCTGCT 1239
QY 1186 CATGAATTTGCTGGGCTACCGCCAGCTGATGAGATGAGAGAGAGAGAGAGAGAG 1245
DB 1240 CATGAATTTGCTGGGCTACCGCCAGCTGATGAGATGAGAGAGAGAGAGAGAGAG 1299
QY 1246 GGAATCTTCTGCTACCTGAGATGCTGCTGAGCAATTCACATGAGGCTTGAAGAGCTT 1305
DB 1300 GGAATCTTCTGCTACCTGAGATGCTGCTGAGCAATTCACATGAGGCTTGAAGAGCTT 1359
QY 1306 TGCTGCACTGAGTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1365
DB 1360 TGCTGCACTGAGTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1419
QY 1366 TCTGTGAGCTTAACTACATGCTGTGGGATCAACACTGATGAGTGTGCTCCACAG 1425
DB 1420 TCTGTGAGCTTAACTACATGCTGTGGGATCAACACTGATGAGTGTGCTCCACAG 1479
QY 1426 TGCTGAGCAGAGAGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1485
DB 1480 TGCTGAGCAGAGAGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1539
QY 1486 TATGTGAGCAGACATCCACAGAGTGAAGAGGATGCTGCTGCTGCTGCTGCTGCT 1545
DB 1540 TATGTGAGCAGACATCCACAGAGTGAAGAGGATGCTGCTGCTGCTGCTGCTGCT 1599

QY 1546 TCTGTGAGCAGACCTTCAAGACCTTGTGAGGCTGAGGCTTATTAAGACAGACAGT 1605
DB 1600 TCTGTGAGCAGACCTTCAAGACCTTGTGAGGCTGAGGCTTATTAAGACAGACAGT 1659
QY 1606 ATCAGTGAATTTGATTCATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1665
DB 1660 ATCAGTGAATTTGATTCATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1719
QY 1666 TTCACCAAGAGAGCTCAGAGCTTTCACAGAGATGC 1703
DB 1720 TTCACCAAGAGAGCTCAGAGCTTTCACAGAGATGC 1757

RESULT 3

US-09-998-2595
; Sequence 2595, Application US/09998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolk, John A.
; APPLICANT: Xu, Jlangchun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Meagher, Madeleine Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.561
; CURRENT APPLICATION NUMBER: US/09/998,598
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corlax Invention Disclosure Database
; SEQ ID NO 2595
; LENGTH: 505
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-998-598-2595

Query Match 19.7%; Score 403; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 8.5e-203;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGGTGTGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1360
DB 1 GGGTGTGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 60
QY 1361 CTACATCTCTAGGCTTAACTACATGCTGTGGGATGACACTGAGTGTGCTGCTC 1420
DB 61 CTACATCTCTAGGCTTAACTACATGCTGTGGGATGACACTGAGTGTGCTGCTC 120
QY 1421 ACAGTGTCAAGCAGAGAGACTTTGTGCTCATGCTGCTGCTGCTGCTGCTGCTGCT 1480
DB 121 ACAGTGTCAAGCAGAGAGACTTTGTGCTCATGCTGCTGCTGCTGCTGCTGCTGCT 180
QY 1481 AACCTATGAG 1540
DB 181 AACCTATGAG 240
QY 1541 GATCTTCTGCTGTGGGAGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1600
DB 241 GATCTTCTGCTGTGGGAGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 300
QY 1601 ACAGTATGAG 1660
DB 301 ACAGTATGAG 360
QY 1661 GATCTTCAACCAAG 1703
DB 361 GATCTTCAACCAAG 403

RESULT 4

US-09-919-580-194
; Sequence 194, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:

APPLICANT: Pyle, Ruth
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.552
CURRENT APPLICATION NUMBER: US/09/919,580
CURRENT FILING DATE: 2001-07-30
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 194
LENGTH: 517
TYPE: DNA
ORGANISM: Homo sapiens
US-09-919-580-194

Query Match 19.7%; Score 403; DB 10; Length 517;
Best Local Similarity 100.0%; Pred. No. 8.5e-203; Indels 0; Gaps 0;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGGTTGGCCACCGTGGTGTGAGCCTCAGTCACTTCTGTGAATGCTTGTGAGCCTTGC 1360
DB 13 GGGTTGGCCACCGTGGTGTGAGCCTCAGTCACTTCTGTGAATGCTTGTGAGCCTTGC 72
QY 1361 CTACATCTGAGCCTTAAGTATGATGCTGTGGTATACACAGTGTGATGCTTC 1420
DB 73 CTACATCTGAGCCTTAAGTATGATGCTGTGGTATACACAGTGTGATGCTTC 132
QY 1421 ACACGTGCTCAGCAGAGAGCTTTGTGTCATGCTGTGTAGAAAACAGCTGGGG 1480
DB 133 ACACGTGCTCAGCAGAGAGCTTTGTGTCATGCTGTGTAGAAAACAGCTGGGG 192
QY 1481 AACCTATGTGACAGACATCCACAGTGAAGAGGATTTGCTTCTTCTTCTT 1540
DB 193 AACCTATGTGACAGACATCCACAGTGAAGAGGATTTGCTTCTTCTTCTT 252
QY 1541 GATCTTCGTGCTGGGAGAGCTCAGAGACTTGTGGCCGGAGGCTATTAGCAGCAG 1600
DB 253 GATCTTCGTGCTGGGAGAGCTCAGAGACTTGTGGCCGGAGGCTATTAGCAGCAG 312
QY 1601 ACAGTATCAGTGAATGATCCATAAACCCTGCTCCACATCTTGGCCATGGGAGAT 1660
DB 313 ACAGTATCAGTGAATGATCCATAAACCCTGCTCCACATCTTGGCCATGGGAGAT 372
QY 1661 GATCTTCACCAAGAGCTCAGCAGATTTTCCACAGAGATGC 1703
DB 373 GATCTTCACCAAGAGCTCAGCAGATTTTCCACAGAGATGC 415

RESULT 5
US-09-919-580-264
Sequence 264, Application US/09919580
Patent No. US20020110832A1
GENERAL INFORMATION:
APPLICANT: Pyle, Ruth
APPLICANT: Xu, Jiangchun
APPLICANT: Secrist, Heather
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.552
CURRENT APPLICATION NUMBER: US/09/919,580
CURRENT FILING DATE: 2001-07-30
NUMBER OF SEQ ID NOS: 934
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 264
LENGTH: 389
TYPE: DNA
ORGANISM: Homo sapiens
US-09-919-580-264

Query Match 5.6%; Score 114; DB 10; Length 389;
Best Local Similarity 100.0%; Pred. No. 1.3e-49; Indels 0; Gaps 0;
Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGGCATCTGTATATACTAAATCCCAATTAAGTTCGTGTAGATGTCCCTTTTATG 1989
DB 266 TGGCATCTGTATATACTAAATCCCAATTAAGTTCGTGTAGATGTCCCTTTTATG 325
QY 1990 CTCTTAATTAATTAAGAGTAATGTTCAATTTTATGGATCTCAAAAAA 2043
DB 326 CTCTTAATTAATTAAGAGTAATGTTCAATTTTATGGATCTCAAAAAA 379

RESULT 6
US-09-927-602-1
Sequence 1, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
TITLE OF INVENTION: Dystrophy
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2544
TYPE: DNA
ORGANISM: Homo Sapien
NAME/KEY: CDS
LOCATION: (693)...(1877)
US-09-927-602-1

Query Match 2.9%; Score 59; DB 9; Length 2544;
Best Local Similarity 100.0%; Pred. No. 1.9e-20; Indels 0; Gaps 0;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCGCTCCTACAGCAGCAGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGAC 682
DB 1180 AGGCTGCGCGCTCCTACAGCAGCAGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGAC 1238

RESULT 7
US-09-927-602-38
Sequence 38, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
TITLE OF INVENTION: Dystrophy
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 38
LENGTH: 48436
TYPE: DNA
ORGANISM: Homo Sapien
US-09-927-602-38

Query Match 2.9%; Score 59; DB 9; Length 48436;
Best Local Similarity 100.0%; Pred. No. 1.9e-20; Indels 0; Gaps 0;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
DB 48024 AGGCTGCGCGCTCCTACAGCAGCAGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGAC 48082

RESULT 8
US-09-867-701-9458
; Sequence 9458, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9458
; LENGTH: 167
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9458

Query Match 1.1%; Score 22; DB 10; Length 167;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 GAAGACGACAGAGGTTAGAG 44
Db 3 GAAGACGACAGAGGTTAGAG 24

RESULT 9
US-09-867-701-9067
; Sequence 9067, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9067
; LENGTH: 171
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9067

Query Match 1.1%; Score 22; DB 10; Length 171;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 GAAGACGACAGAGGTTAGAG 44
Db 9 GAAGACGACAGAGGTTAGAG 30

RESULT 10
US-09-867-701-9185
; Sequence 9185, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497

; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9185
; LENGTH: 187
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9185

Query Match 1.1%; Score 22; DB 10; Length 187;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 GAAGACGACAGAGGTTAGAG 44
Db 32 GAAGACGACAGAGGTTAGAG 53

RESULT 11
US-09-867-701-9284
; Sequence 9284, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9284
; LENGTH: 191
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9284

Query Match 1.1%; Score 22; DB 10; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 23 GAAGACGACAGAGGTTAGAG 44
Db 29 GAAGACGACAGAGGTTAGAG 50

RESULT 12
US-09-960-352-11099/C
; Sequence 11099, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengbing
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11099
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-BOVMS1-008-Q1-E1-DB
US-09-960-352-11099

Query Match 1.0%; Score 20; DB 10; Length 398;
Best Local Similarity 100.0%; Pred. No. 9;

Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GATGCCATCTGGCTCTAT 194
|||||
Db 240 GATGCCATCTGGCTCTAT 221

RESULT 13
US-10-027-632-140660
; Sequence 140660, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Mang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140660
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-140660

Query Match 1.0%; Score 20; DB 13; Length 570;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 GAAAGACCCCTCCCTCAACC 719
|||||
Db 273 GAAAGACCCCTCCCTCAACC 292

RESULT 14
US-10-205-823-334/C
; Sequence 334, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannevarapu, Kanjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangl
; APPLICANT: Monney, Angela M.
; APPLICANT: Glatte, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356

; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 334
; LENGTH: 11102
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-823-334

Query Match 1.0%; Score 20; DB 14; Length 11102;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1523 TGCTCTCTCTCTCTCTGA 1542
|||||
Db 722 TGCTCTCTCTCTCTCTGA 703

RESULT 15
US-10-003-806-6
; Sequence 6, Application US/10003806
; Publication No. US20020119929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Collin E.
; APPLICANT: Agoulnik, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P02066US1/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
US-10-003-806-6

Query Match 1.0%; Score 20; DB 13; Length 180557;
Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 20: Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1893 TTTTCCATTACATAGAAA 1912
|||||
Db 2567 TTTTCCATTACATAGAAA 2586

Search completed: August 15, 2003, 04:41:29
Job time : 1507 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2003, 03:08:38 ; Search time 7281 Seconds

11478.968 Million cell updates/sec

Title: US-09-645-078-1
 Page: 3043

Sequence: 1 gaattccattgttggta.....tgggatcctaataaaaaaaa 2043

Scoring table: OLIGO_NUC

Searched: 2888711 seqs, 20454813386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 3519286

Minimum DB seq length: 300 /
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database

1:	gb_ba:*
2:	gb_htg:*
3:	gb_in:*
4:	gb_om:*
5:	gb_ov:*
6:	gb_pac:*
7:	gb_ph:*
8:	gb_pl:*
9:	gb_pr:*
10:	gb_ro:*
11:	gb_sbs:*
12:	gb_sy:*
13:	gb_un:*
14:	gb_vl:*
15:	em_ba:*
16:	em_fun:*
17:	em_hum:*
18:	em_in:*
19:	em_mu:*
20:	em_om:*
21:	em_or:*
22:	em_ov:*
23:	em_pat:*
24:	em_ph:*
25:	em_pl:*
26:	em_sts:*
27:	em_un:*
28:	em_un:*
29:	em_vl:*
30:	em_hcg_hum:*
31:	em_hcg_in:*
32:	em_hcg_other:*
33:	em_hcg_mus:*
34:	em_hcg_pln:*
35:	em_hcg_rtd:*
36:	em_hcg_mem:*
37:	em_hcg_yrt:*
38:	em_sy:*
39:	em_htgo_hum:*
40:	em_htgo_mus:*
41:	em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1722	84.3	1992	9	AF280088	AF280088 Homo sapi
2	1653	80.9	2037	9	BC035282	BC035282 Homo sapi
3	1598	78.2	2021	9	AK026635	AK026635 Homo sapi
4	1597	78.2	183328	9	AC010547	AC010547 Homo sapi
5	1597	78.2	206943	2	AC138848	AC138848 Homo sapi
6	1547	75.7	1979	6	BD127258	BD127258 Homo sapi
7	1547	75.7	1979	6	AK074746	AK074746 Homo sapi
8	1547	75.7	2032	6	AR203335	AR203335 Sequence
9	1547	75.7	2032	6	BD134772	BD134772 Glycosyls
10	1547	75.7	2032	9	AF131235	AF131235 Homo sapi
11	1262	61.8	1333	9	AF149783	AF149783 Homo sapi
12	1155	56.4	2065	6	AR310485	AR310485 Sequence
13	695	34.0	877	6	BD124832	BD124832 Primer to
14	695	34.0	877	6	BD126950	BD126950 Primer to
15	403	19.7	517	6	AX381256	AX381256 Sequence
16	114	5.6	389	6	AX381326	AX381326 Sequence
17	105	5.1	548	6	BD125617	BD125617 Primer fo
18	59	2.9	1462	9	AF176839	AF176839 Homo sapi
19	59	2.9	1647	6	AX327330	AX327330 Sequence
20	59	2.9	2170	9	AF176838	AF176838 Homo sapi
21	59	2.9	2544	9	AF179990	AF179990 Homo sapi
22	59	2.9	3378	9	AF246718	AF246718 Homo sapi
23	59	2.9	3786	9	AF280086	AF280086 Homo sapi
24	59	2.9	71503	9	AF219991	AF219991 Homo sapi
25	59	2.9	157358	9	AC009163	AC009163 Homo sapi
26	59	2.9	194832	9	AC025287	AC025287 Homo sapi
27	59	2.9	208185	6	AC009105	AC009105 Homo sapi
28	27	1.3	1926	6	BD134773	BD134773 Glycosyls
29	27	1.3	1926	6	AF131236	AF131236 Mus muscu
30	27	1.3	2201	10	AF109155	AF109155 Mus muscu
31	27	1.3	232745	2	AC130174	AC130174 Rattus no
32	27	1.3	262679	2	AC110314	AC110314 Rattus no
33	24	1.2	228569	2	AC121740	AC121740 Rattus no
34	24	1.2	235173	2	AC133449	AC133449 Rattus no
35	24	1.2	240901	2	AC096332	AC096332 Rattus no
36	24	1.2	242253	2	AC095651	AC095651 Rattus no
37	24	1.2	246877	2	AC137365	AC137365 Rattus no
38	24	1.2	258097	2	AC129764	AC129764 Rattus no
39	24	1.2	271330	2	AC094919	AC094919 Rattus no
40	24	1.1	1740	10	AF176841	AF176841 Mus muscu
41	23	1.1	1899	10	AF176840	AF176840 Mus muscu
42	23	1.1	110000	2	AC099262	AC099262 Continuation (2 of
43	23	1.1	159170	2	AC115914	AC115914 Mus muscu
44	23	1.1	192365	2	AC138691	AC138691 Rattus no
45	23	1.1	194439	2	AC109022	AC109022 Rattus no

ALIGNMENTS

RESULT 1	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE	AUTHORS
AF280088	AF280088	1992 bp mRNA linear PRI 20-FEB-2001	Homo sapiens L-selectin ligand sulfoltransferase GST-3 mRNA, complete cds.	AF280088	AF280088.1 GI:12060807	Homo sapiens (human)	Homo sapiens	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo. 1 (bases 1 to 1992)	Hemmetich, S., Lee, J.K., Bhakta, S., Bistrup, A., Ruddle, N.R. and Rosen, S.D.

TITLE Chromosomal localization and genomic organization for the galactose/ N-acetylglucosamine/ N-acetylglucosamine 6-O-sulfotransferase gene family
JOURNAL Glycobiology 11 (1), 75-87 (2001)
MEDLINE 21096027
PUBMED 11181564
REFERENCE 2 (bases 1 to 1992)
AUTHORS Hemmerich, S., Bhakta, S., Lee, J.-K., Bistrup, A., Ruddie, N.R. and Rosen, S.D.
TITLE Direct Submision
JOURNAL Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience, 3401 Hillview Ave., Palo Alto, CA 94304, USA
FEATURES Location/Qualifiers
source 1. 1992
organism "Homo sapiens"
mol_type "mRNA"
db_xref "taxon:9606"
chromosome "16"
map "16q23.1-q23.2"
cell_type "high endothelial"
tissue_type "tonsil"
gene 1. 1992
gene "GST3"
CDS 122..1282
note "similar to Homo sapiens sequence encoded by GenBank accession Number AF131235"
product "L-selectin ligand sulfotransferase GST-3"
codon_start 1
protein_id "A048246.1"
db_xref "GI:12060808"
translation "MLPRKMKLLFLVSQMATLAFPHMYSHNISLSMKAPERKH VLVLSMRSGSSVVGOLFQHPDVTYLMERAMHWMTFKOSTAMLMHVRDILRAVE LCDMSYFDVMEGRPRSSLEFQWENSRLCSAPACDIIIPQEDIIIPRACRLCSQOP FEWEKASYSYSHVLEKFEMLQSLYDLKPSLNLHIVHLRPRPRAVFRSRETRY GDLMDISRIYMGHOKLEKKEDOPYVVMYVICOSELEIKTIGSLPRALOEKRLVRY EDLAPRVAOTSRMYEFVGLIEFLPHLOTWHNTTREGKGDHAFHNADALNVQAM RMSLPKXRSRQKAGDAMNLTGRHVSQEQRLDLDLSTWTPVPIQH"
BASE COUNT 482 a 540 c 489 g 481 t
ORIGIN
Query Match 84.3%; Score 1722; DB 9; Length 1992;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1722; Conservative 0; **Mismatches** 0; **Indels** 0; **Gaps** 0;

427 CGTCTTCTGTGGACATGAGCGCTCTTTGATGCCCTACAGTAAGCACTGGTCCCGGAGACA 486
 |||||||
 421 CGTCTTCTGTGGACATGAGCGCTCTTTGATGCCCTACAGTAAGCACTGGTCCCGGAGACA 480
 |||||||
 487 GTCCAGCGCTCTTACAGGGAGAAACCGGGGCGCTGTGTTGCAACCTGCCGTGCAT 546
 |||||||
 481 GTCCAGCGCTCTTACAGGGAGAAACCGGGGCGCTGTGTTGCAACCTGCCGTGCAT 540
 |||||||
 547 CATCCCAAGATGAAATCATCCCGGGGCTCACTGCAGGCTCCTGTGCAGTCAAGGCC 606
 |||||||
 541 CATCCCAAGATGAAATCATCCCGGGGCTCACTGCAGGCTCCTGTGCAGTCAAGGCC 600
 |||||||
 607 CTTTGAAGTGTGGAGAAAGGCTCCGCTCTACAGCCACGTGTGTCTCAAGAGTGGC 666
 |||||||
 601 CTTTGAAGTGTGGAGAAAGGCTCCGCTCTACAGCCACGTGTGTCTCAAGAGTGGC 660
 |||||||
 667 CTTCTTCAACCTGACAGTCCCTTACCGCGCTGTGAAGAACCCCTCCCTCAACCTGCATAT 726
 |||||||
 661 CTTCTTCAACCTGACAGTCCCTTACCGCGCTGTGAAGAACCCCTCCCTCAACCTGCATAT 720
 |||||||
 727 CGTGCACCTGTGTCGGGACCCCGGGGCGCTGTCCCTTCCGGAAGACGCAAAAGGAGA 786
 |||||||
 721 CGTGCACCTGTGTCGGGACCCCGGGGCGCTGTCCCTTCCGGAAGACGCAAAAGGAGA 780
 |||||||
 787 TCTCATGATTTGACAGTCGCAATTTGTAATGGGGGAGCAGTACGCAAAATCTCAAGAGAGA 846
 |||||||
 781 TCTCATGATTTGACAGTCGCAATTTGTAATGGGGGAGCAGTACGCAAAATCTCAAGAGAGA 840
 |||||||
 847 CCAACCTTATGATGATGAGAGTGCATCTGTCCTCAAGCCACAGCTGAGATCTACAAGACAT 906
 |||||||
 841 CCAACCTTATGATGATGAGAGTGCATCTGTCCTCAAGCCACAGCTGAGATCTACAAGACAT 900
 |||||||
 907 CCAATCTTGTCCCAAGGCGCTGTGAGAAAGCCTACTGCTGTGCGCTATGAGAGACCTGGC 966
 |||||||
 901 CCAATCTTGTCCCAAGGCGCTGTGAGAAAGCCTACTGCTGTGCGCTATGAGAGACCTGGC 960
 |||||||
 967 TCGAGCCCTGTGTCGGGACCAAGCTTCCGAATGTGTAATTCGTGGAGTGGAAATCTTGGC 1026
 |||||||
 961 TCGAGCCCTGTGTCGGGACCAAGCTTCCGAATGTGTAATTCGTGGAGTGGAAATCTTGGC 1020
 |||||||
 1027 CCATCTTACAGACCTGGGTGCATTAACATCAACCCGAGGCAAGGGCATGGTACACGCTTT 1086
 |||||||
 1021 CCATCTTACAGACCTGGGTGCATTAACATCAACCCGAGGCAAGGGCATGGTACACGCTTT 1080
 |||||||
 1087 CCAACAAATGCCAGGATGCCCTTAATGTCTCCACAGGCTTGGCGCTGTGTCCTTA 1146
 |||||||
 1081 CCAACAAATGCCAGGATGCCCTTAATGTCTCCACAGGCTTGGCGCTGTGTCCTTA 1140
 |||||||
 1147 TGAAGGCTTTCGCACTTCAGAAAGCCTGTGGCGAATGCCAATTTGCTGGGCTACACG 1206
 |||||||
 1141 TGAAGGCTTTCGCACTTCAGAAAGCCTGTGGCGAATGCCAATTTGCTGGGCTACACG 1200
 |||||||
 1207 CCAAGTCAGATCTGAACAAGAACAGAAACCTGTGTGCTGATCTTCTGTCTACCTGGAC 1266
 |||||||
 1201 CCAAGTCAGATCTGAACAAGAACAGAAACCTGTGTGCTGATCTTCTGTCTACCTGGAC 1260
 |||||||
 1267 TGTCCCTGAGCAAAATCCACTAAGAGGCTTGAGAAAGCTTGTGCTCCACCTGTGTACGCC 1326
 |||||||
 1261 TGTCCCTGAGCAAAATCCACTAAGAGGCTTGAGAAAGCTTGTGCTCCACCTGTGTACGCC 1320
 |||||||
 1327 TCACTACCTTTCCTGAATGCTTCTGAGCCTTGCCTGCAATCTGTGAGCCCTTAATCTACAG 1386
 |||||||
 1321 TCACTACCTTTCCTGAATGCTTCTGAGCCTTGCCTGCAATCTGTGAGCCCTTAATCTACAG 1380
 |||||||
 1387 TCTGTGGGTATCACTAGAGTGTGAATGTGTCTCCACAGCTGTCTCAAGAGAGACTTTT 1446
 |||||||
 1381 TCTGTGGGTATCACTAGAGTGTGAATGTGTCTCCACAGCTGTCTCAAGAGAGACTTTT 1440
 |||||||
 1447 GTGTCCATGCTTGTGTGTAAGAAACAGACTGGGGAACCTTAATGTGAGACAGACATCCAC 1506
 |||||||
 1441 GTGTCCATGCTTGTGTGTAAGAAACAGACTGGGGAACCTTAATGTGAGACAGACATCCAC 1500
 |||||||
 1507 CAGTGAAGAGGATATGCTCTTCTTTCTGTGATCTTCTGCTGTGCGGAGACATTCAG 1566

QY 951 GCTATAGAGACCTGGCTCGAGCCCTGTGGCCCGACACTTCCGAAATGATGAATTCGTGG 1010
 Db 901 GCTATAGAGACCTGGCTCGAGCCCTGTGGCCCGACACTTCCGAAATGATGAATTCGTGG 960
 QY 1011 GATTGGAATTCCTGGCCCATCTTCAGACCTGGGTGCATATACACCCCGGGAAGGCA 1070
 Db 961 GATTGGAATTCCTGGCCCATCTTCAGACCTGGGTGCATATACACCCCGGGAAGGCA 1020
 QY 1071 TGGGTGACACGAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGC 1130
 Db 1021 TGGGTGACACGAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGC 1080
 QY 1131 GCTGTCTTGTGGCCCTTGAAGAGCTTTCGACTTGAAGAGCTTGGCGATGCCATGA 1190
 Db 1081 GCTGTCTTGTGGCCCTTGAAGAGCTTTCGACTTGAAGAGCTTGGCGATGCCATGA 1140
 QY 1191 ATTGTGTGGGCTTCGCGCCAGCTGAGATGAGATGAACAAACGAGAAACCTTCTGTGATC 1250
 Db 1141 ATTGTGTGGGCTTCGCGCCAGCTGAGATGAGATGAACAAACGAGAAACCTTCTGTGATC 1200
 QY 1251 TTCTGTCTACCTGGAGCTGTCCCTGAGCAAAATCCACTAAGAGGCTTGAAGAGCTTGTG 1310
 Db 1201 TTCTGTCTACCTGGAGCTGTCCCTGAGCAAAATCCACTAAGAGGCTTGAAGAGCTTGTG 1260
 QY 1311 CCACCTGGGTGACGCTCAGCTCACTTCTCTGTAATGCTTCTGAGCCTTGCCTACATCT 1370
 Db 1261 CCACCTGGGTGACGCTCAGCTCACTTCTCTGTAATGCTTCTGAGCCTTGCCTACATCT 1320
 QY 1371 GAGCCTTAACATGCTGTGGGTATGACACAGTGAAGAGTGTGTCACAGCTGCTC 1430
 Db 1321 GAGCCTTAACATGCTGTGGGTATGACACAGTGAAGAGTGTGTCACAGCTGCTC 1380
 QY 1431 AAGCAGAGAGACTTTTGTGTGTCATGCTTGTCTAGAAAAACAGACTGGGGAACCTTATGT 1490
 Db 1381 AAGCAGAGAGACTTTTGTGTGTCATGCTTGTCTAGAAAAACAGACTGGGGAACCTTATGT 1440
 QY 1491 GAGCAGACATCCACACAGTGAAGAGGATGTGCTTCTTCTTCTTGTGATCTTCTG 1550
 Db 1441 GAGCAGACATCCACACAGTGAAGAGGATGTGCTTCTTCTTCTTGTGATCTTCTG 1500
 QY 1551 TCTGGCAGACTTCAGAGACTTGTGGCCTGAGAGGCTTATTAAGCAGACAGATATCAG 1610
 Db 1501 TCTGGCAGACTTCAGAGACTTGTGGCCTGAGAGGCTTATTAAGCAGACAGATATCAG 1560
 QY 1611 TGAATTTGATCCATAAAGCTCCCTGTCCACATCTTCCCAATGGGGAATGATCTTTCAC 1670
 Db 1561 TGAATTTGATCCATAAAGCTCCCTGTCCACATCTTCCCAATGGGGAATGATCTTTCAC 1620
 QY 1671 CAAGAGCTCACAGCACTTTCACAGAGATGC 1703
 Db 1621 CAAGAGCTCACAGCACTTTCACAGAGATGC 1653
 RESULT 3
 AK026635 2011 bp mRNA linear PRI 29-SEP-2000
 LOCUS Homo sapiens cDNA: FLJ22982 fls. clone KAT11454, highly similar to
 DEFINITION AF131235 Homo sapiens N-acetylglyucosamine 6-O-sulfotransferase
 mRNA.
 ACCESSION AK026635
 VERSION AK026635.1 GI:10439531
 KEYWORDS oligo capping, fls (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (sites)
 Matanabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H.,
 Ota, T., Suzuki, Y., Odayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
 Nakamura, Y., Isogai, T. and Sugano, S.
 NEDO human cDNA sequencing project
 TITLE NEDO human cDNA sequencing project
 JOURNAL Unpublished
 REFERENCE 2 (bases 1 to 2011)

AUTHORS Sugano, S., Suzuki, Y., Ota, T., Odayashi, M., Nishi, T., Isogai, T.,
 Shibahara, T., Tanaka, T. and Nakamura, Y.
 TITLE Direct Submission
 JOURNAL Submitted (29-NOV-2000) Sumio Sugano, Institute of Medical Science,
 University of Tokyo, Laboratory of Genome Structure Analysis, Human
 Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
 Japan (E-mail: cdna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,
 Fax: 81-3-5449-5416)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of
 International Trade and Industry of Japan: cDNA full insert
 sequencing: Research Association for Biotechnology: cDNA library
 construction, 5'-6' end one pass sequencing: Department of
 Virology and Human Genome Center, Institute of Medical Science,
 University of Tokyo (partly supported by Science and Technology
 Agency).
 FEATURES
 source location/Qualifiers
 1. 2011
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="KAT11454"
 /cell_line="KATO III"
 /cell_type="signet-ring cell carcinoma"
 /note="cloning vector pME18SFL3"
 /note="highly similar to AF131235 Homo sapiens
 N-acetylglyucosamine 6-O-sulfotransferase mRNA"
 ORIGIN
 BASE COUNT 482 a 541 c 475 g 513 t
 Query Match 78.2%; Score 1598; DB 9; Length 2011;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1598; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 106 CAAGGCTTCACATTCAGACATGCTAGCTTAAATAAAGTCTGCTGTTTCT 165
 Db 50 CAAGGCTTCACATTCAGACATGCTAGCTTAAATAAAGTCTGCTGTTTCT 109
 QY 166 GATTTCAGATGGCATCTTGCTATCTTCCACATATAGCAGCACAAATCAGATC 225
 Db 110 GATTTCAGATGGCATCTTGCTATCTTCCACATATAGCAGCACAAATCAGATC 169
 QY 226 CCTGTATGAAGGACAGCCGAGCGATGACGTGTGCTTCTTCTGCGCTC 285
 Db 170 CCTGTATGAAGGACAGCCGAGCGATGACGTGTGCTTCTTCTGCGCGCTC 229
 QY 286 TGCGCTCTCTTGTGGGAGCTTTTGGGAGCAACCAAGTCTTCTACTGATGGA 345
 Db 230 TGCGCTCTCTTGTGGGAGCTTTTGGGAGCAACCAAGTCTTCTACTGATGGA 289
 QY 346 GCCCGCTGACAGTGTGATGACCTTCAAGCAGACACCGCTGATGCTGCATGCG 405
 Db 290 GCCCGCTGACAGTGTGATGACCTTCAAGCAGACACCGCTGATGCTGCATGCG 349
 QY 406 TGTGCGGAGTGTATACGGGCGCTCTTCTTGTGCGACATGAGCGCTTTGATGCTATAT 465
 Db 350 TGTGCGGAGTGTATACGGGCGCTCTTCTTGTGCGACATGAGCGCTTTGATGCTATAT 409
 QY 466 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTCAAGTGGGAGAAAGCGCGGCGCTG 525
 Db 410 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTCAAGTGGGAGAAAGCGCGGCGCTG 469
 QY 526 TTCTGACCTGCTGTGATCATATCCACAAAGATGAATATATCCCGGGGCTCACTGAG 585
 Db 470 TTCTGACCTGCTGTGATCATATCCACAAAGATGAATATATCCCGGGGCTCACTGAG 529
 QY 586 GCTCTGTGAGTCAACAGCCCTTTGAGGTGTGAGAAAGGCGCTGCGCTCTTACAGCA 645
 Db 530 GCTCTGTGAGTCAACAGCCCTTTGAGGTGTGAGAAAGGCGCTGCGCTCTTACAGCA 589
 QY 646 CGTGTGCTCAAGAGAGTGGCTTTCACATCGACGTCCCTTACCCTGTGTGAAGA 705
 Db 646 CGTGTGCTCAAGAGAGTGGCTTTCACATCGACGTCCCTTACCCTGTGTGAAGA 705

```

Db 590 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTACCCGCTGTGAAGA 649
Oy 706 CCCCTCCCTCAACCTGCATATCGACACCTGGTCCGGAGACCCCGGGCCGTGTCCGTC 765
Db 650 CCCCTCCCTCAACCTGCATATCGACACCTGGTCCGGAGACCCCGGGCCGTGTCCGTC 709
Oy 766 CCGAAGAGCACAAGAGAGATCTCATATGATGACAGTGCATTTGTGATGGGCGACATGA 825
Db 710 CCGAAGAGCACAAGAGAGATCTCATATGATGACAGTGCATTTGTGATGGGCGACATGA 769
Oy 826 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTATGACAGTCACTCCCAAGCCA 885
Db 770 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTATGACAGTCACTCCCAAGCCA 829
Oy 886 GCTGAGATCTACAAAGACATCCAGTCTCTGCCAAGGCGCTGAGAGAGAGCTACCTGCT 945
Db 830 GCTGAGATCTACAAAGACATCCAGTCTCTGCCAAGGCGCTGAGAGAGAGCTACCTGCT 889
Oy 946 TGTGCGGTATGAGAGACCTGGCTCGAGCCCTGTGGCCAGACTTCCGAAATGTATGAAT 1005
Db 890 TGTGCGGTATGAGAGACCTGGCTCGAGCCCTGTGGCCAGACTTCCGAAATGTATGAAT 949
Oy 1006 CGTGGGATTTGAATTTTGGCCCATTTTCAGACCTGGGTGATTAACATACCCGAGCAA 1065
Db 950 CGTGGGATTTGAATTTTGGCCCATTTTCAGACCTGGGTGATTAACATACCCGAGCAA 1009
Oy 1066 GGGCATGGGTGACACAGCTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGC 1125
Db 1010 GGGCATGGGTGACACAGCTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGC 1069
Oy 1126 TTGGCGGTGTCTTGGCCCATTTTCAGAAAGTTTTCAGACTTCAGAAAGCTGTGGCGATG 1185
Db 1070 TTGGCGGTGTCTTGGCCCATTTTCAGAAAGTTTTCAGACTTCAGAAAGCTGTGGCGATG 1129
Oy 1186 CATGAATTTGTGGGCTTACCCGACAGCTCAGATCTGAACAAAGAGAGAAACCTGTGCT 1245
Db 1130 CATGAATTTGTGGGCTTACCCGACAGCTCAGATCTGAACAAAGAGAGAAACCTGTGCT 1189
Oy 1246 GGATCTTCTGTCTACCTGAGATGTCTGAGCAATCCACTAAGAGGGTTAGAGAGCTT 1305
Db 1190 GGATCTTCTGTCTACCTGAGATGTCTGAGCAATCCACTAAGAGGGTTAGAGAGCTT 1249
Oy 1306 TGTGACCACTGTGTGACGCTCAGTCACTTCTCTGAATGCTTCTGACCTTGCCTACA 1365
Db 1250 TGTGACCACTGTGTGACGCTCAGTCACTTCTCTGAATGCTTCTGACCTTGCCTACA 1309
Oy 1366 TCTCTGAGCCTTACTACATGTCTGTGGTATCAACACTGATGTGATGTCTCCACAG 1425
Db 1310 TCTCTGAGCCTTACTACATGTCTGTGGTATCAACACTGATGTGATGTCTCCACAG 1369
Oy 1426 TGTCAAGCACAAGAGACTTGTGTCCATGCTGTGTCTAGAAACAGAGCTGGGGAACCT 1485
Db 1370 TGTCAAGCACAAGAGACTTGTGTCCATGCTGTGTCTAGAAACAGAGCTGGGGAACCT 1429
Oy 1486 TATGTGAGCAGACATCCACAGTGAAGAGGGTATTTGCTTCTTCTTCTGATCT 1545
Db 1430 TATGTGAGCAGACATCCACAGTGAAGAGGGTATTTGCTTCTTCTTCTGATCT 1489
Oy 1546 TCTGTCTGGGCAAGCTTCAAGACTTTGTGGCTCGAGGCGCTATTAAGCAGACAGCT 1605
Db 1490 TCTGTCTGGGCAAGCTTCAAGACTTTGTGGCTCGAGGCGCTATTAAGCAGACAGCT 1549
Oy 1606 ATCAGTGAATTTGAATTAAGCTCCCTGTCCACATCTGCCAATGGGGAATGAGACT 1665
Db 1550 ATCAGTGAATTTGAATTAAGCTCCCTGTCCACATCTGCCAATGGGGAATGAGACT 1609
Oy 1666 TTCACAAAGAGCTCACCAGCATTTTCCACAGAGATGC 1703
Db 1610 TTCACAAAGAGCTCACCAGCATTTTCCACAGAGATGC 1647

```

RESULT 4
AC010547

```

LOCUS AC010547 183228 bp DNA linear PRI 29-SEP-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-510M2, complete sequence.
ACCESSION AC010547
VERSION AC010547.9 GI:15808510
KEYWORDS HTG.
ORGANISM Homo sapiens (human)
SOURCE Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 183228)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 183228)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission.
JOURNAL Direct Submission
REFERENCE 3 (bases 1 to 183228)
AUTHORS DOE Joint Genome Institute.
TITLE Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Sep 29, 2001 this sequence version replaced gi:14589436.
COMMENT Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
Genbank flat file format but are available as part
of this entry's ASN.1 file.
-----
FEATURES
source
1. 183228
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-510M2"
BASE COUNT 51901 a 42098 c 40026 g 49203 t
ORIGIN
Query Match 78.2%; Score 1597; DB 9; Length 183228;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 107 AAGCTCTTCACCTTCAGCACAATGCTACTGCTTAAAAAATGAAGCTCTCTGTTCTG 166
Db 142584 AAGGCTTCACCTTCAGCACAATGCTACTGCTTAAAAAATGAAGCTCTCTGTTCTG 142643
Oy 167 GTTTCGCCGATGGGCATTTTGGCTCTATTTCCACATGTACAGCCACAATACAGCTCC 226
Db 142644 GTTTCGCCGATGGGCATTTTGGCTCTATTTCCACATGTACAGCCACAATACAGCTCC 142703
Oy 227 CTGTCTATGAAGGACACAGCCGAGGAGATGACAGTGTGTTCTTCTGCGCGCTCT 286
Db 142704 CTGTCTATGAAGGACACAGCCGAGGAGATGACAGTGTGTTCTTCTGCGCGCTCT 142763
Oy 287 GGCCTCTCTTTTGTGGGCGACCTTTTGGGCGAGCACCAGATGTTTCTACCTGATGGAG 346
Db 142764 GGCCTCTCTTTTGTGGGCGACCTTTTGGGCGAGCACCAGATGTTTCTACCTGATGGAG 142823

```

QY	347	CCCGCGTGGCAGSTGTGGATGACCTTCAAGCAGACACCGCGTGGATGCTGGACATAGGCT	406
Db	142824	CCCGCTTGGCAGCTGTGGATGACCTTCAAGCAGACACCGCGTGGATGCTGGACATAGGCT	142883
QY	407	GTGCGGAGTCGATACGGGCGCTCTCTGTGTGCGCATGAGCGCTCTTGTGATGCTTACATG	466
Db	142884	GTGCGGAGTCGATACGGGCGCTCTCTGTGTGCGCATGAGCGCTCTTGTGATGCTTACATG	142943
QY	467	GAACCTGGTCCCCGGAGACAGTCACGCCCTTTTCAGTGGGAATAACAGCCGGGCGCTGTGT	526
Db	142944	GAACCTGGTCCCCGGAGACAGTCACGCCCTTTTCAGTGGGAATAACAGCCGGGCGCTGTGT	143003
QY	527	TCTGCACTCGCTGTGACATCATCCACAAAGATGAAATATCCCCGGGCTCACTCTACAG	586
Db	143004	TCTGCACTCGCTGTGACATCATCCACAAAGATGAAATATCCCCGGGCTCACTCTACAG	143063
QY	587	CTCCTGTGCACTCAACAGCCCTTTGAGGTGTGGAGAAAGGCGTCCGCTCTACAGCCAC	646
Db	143064	CTCCTGTGCACTCAACAGCCCTTTGAGGTGTGGAGAAAGGCGTCCGCTCTACAGCCAC	143123
QY	647	GTGGGCTCAAGAGAGGTGGCTTCTTCAACCTGCAGTCCCTTACCCGCTGCTGAAGAAG	706
Db	143124	GTGGGCTCAAGAGAGGTGGCTTCTTCAACCTGCAGTCCCTTACCCGCTGCTGAAGAAG	143183
QY	707	CCCTCCCTCAACCTGCAATATCGTGCACCTGTGTCGCGGAGCCCCGGGCGCTGTCTCC	766
Db	143184	CCCTCCCTCAACCTGCAATATCGTGCACCTGTGTCGCGGAGCCCCGGGCGCTGTCTCC	143243
QY	767	CGAGAACCCACAAAGGAGATCTCATGATGTGACAGTCCGATTGTGATGGGGCAGCATGAG	826
Db	143244	CGAGAACCCACAAAGGAGATCTCATGATGTGACAGTCCGATTGTGATGGGGCAGCATGAG	143303
QY	827	CAAAACCTCAAGAAAGAGAGACCACCCCTACTATGTGATGCAGGTCACTGCCCCAAAGCCAG	886
Db	143304	CAAAACCTCAAGAAAGAGAGACCACCCCTACTATGTGATGCAGGTCACTGCCCCAAAGCCAG	143363
QY	887	CTGGAGATCTTACAGACCATCATCGTCTTGGCCCAAGGCCCTGCAGAGAGCTTACCTGCTT	946
Db	143364	CTGGAGATCTTACAGACCATCATCGTCTTGGCCCAAGGCCCTGCAGAGAGCTTACCTGCTT	143423
QY	947	GTGGGCTTNGAGGACCTGGGCTGAGACCCCTGTGGGCCAGACTCCCGAATGTATGAATTC	1006
Db	143424	GTGGGCTTNGAGGACCTGGGCTGAGACCCCTGTGGGCCAGACTCCCGAATGTATGAATTC	143483
QY	1007	GTGGGATTTGAATTTCTTGGCCCATTTGAGACTCTGGGTGCATTAACATCACCCGAGCAAG	1066
Db	143484	GTGGGATTTGAATTTCTTGGCCCATTTGAGACTCTGGGTGCATTAACATCACCCGAGCAAG	143543
QY	1067	GGCATGGGTGACACGCTTTTCCACAAATAATGCCAGGAGATGCCCTTAATGTCTCCAGGCT	1126
Db	143544	GGCATGGGTGACACGCTTTTCCACAAATAATGCCAGGAGATGCCCTTAATGTCTCCAGGCT	143603
QY	1127	TGGGCGTGGTCTTTGGCCCATATGAAAGGTTTCTGCACCTTCAGAAAAGCTGTGGCATATGC	1186
Db	143604	TGGGCGTGGTCTTTGGCCCATATGAAAGGTTTCTGCACCTTCAGAAAAGCTGTGGCATATGC	143663
QY	1187	ATGAATTTGTGGGGCTTACCGCACGTCAGATCTGAACAAGACAGAGAACCTGTGTGCTG	1246
Db	143664	ATGAATTTGTGGGGCTTACCGCACGTCAGATCTGAACAAGACAGAGAACCTGTGTGCTG	143723
QY	1247	GATCTTCTGTCTACCTGAGCTGTGCTTCCCTGAGCAATCCACTTAAGAGGTTGAGAAAGGCTTT	1306
Db	143724	GATCTTCTGTCTACCTGAGCTGTGCTTCCCTGAGCAATCCACTTAAGAGGTTGAGAAAGGCTTT	143783
QY	1307	GCTGCGCACTGGTGTACAGCTTCAGTACCTTCTCTGGAATGCTTCTGAGGCTTGCCTACAT	1366
Db	143784	GCTGCGCACTGGTGTACAGCTTCAGTACCTTCTCTGGAATGCTTCTGAGGCTTGCCTACAT	143843
QY	1367	CTTGAGGCTTAACTACATAGTGTCTGTGGGATATCACTAGTGTGAGATTTGTGTCACATAGCT	1426
Db	143844	CTTGAGGCTTAACTACATAGTGTCTGTGGGATATCACTAGTGTGAGATTTGTGTCACATAGCT	143903

OY	1427	CCTCAGACAGAGACATTTTGTGTCATGCTGTGCTGTGAAACAGACTGGGGAACCTT	1486
Db	143904	GCTCAGACAGAGACATTTTGTGTCATGCTGTGCTGTGAAACAGACTGGGGAACCTT	143965
OY	1487	ATGTGAGACAGCACATCCACCCAGTGAACAGGAGTATGCTCTTCTTCTTTCTTGATCTT	1546
Db	143964	ATGTGAGACAGCACATCCACCCAGTGAACAGGAGTATGCTCTTCTTCTTTCTTGATCTT	144023
OY	1547	CCTGCTGTGGCGACATCTTCAGACATTTGTGGGCTGGAGGCCATTTAAGCAGCAGAGTA	1606
Db	144024	CCTGCTGTGGCGACATCTTCAGACATTTGTGGGCTGGAGGCCATTTAAGCAGCAGAGTA	144088
OY	1607	TCAGTGAATTCATCCATTAACCTCCCTGCCACATCTTCCCAATGGGGAATGATCTT	1666
Db	144084	TCAGTGAATTCATCCATTAACCTCCCTGCCACATCTTCCCAATGGGGAATGATCTT	144143
OY	1667	TCACCAAGAGCTCACACAGCATTTTCCACAGAGATCC	1703
Db	144144	TCACCAAGAGCTCACACAGCATTTTCCACAGAGATCC	144180
RESULT 5			
AC138848/c			
LOCUS	AC138848	206943 bp	DNA
DEFINITION	Human sapiens chromosome 5 clone RP11-1301B21, WORKING DRAFT		
ACCESSION	AC138848		
VERSION	AC138848.1	GI:27805260	
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	1 (bases 1 to 206943)		
TITLE	DOE Joint Genome Institute.		
JOURNAL	Sequencing of Human Chromosome 5		
REFERENCE	Unpublished		
AUTHORS	2 (bases 1 to 206943)		
TITLE	DOE Joint Genome Institute.		
JOURNAL	Direct Submission		
COMMENT	Submitted (21-JAN-2003) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
	-----Genome Center		
	Center: Joint Genome Institute		
	Center Code: JGI		
	Web site: http://www.jgi.doe.gov		

	Project Information		
	Center Project Name: 2719263		
	Center clone name: RPCI-11_1301B21		

	Summary Statistics		
	Consensus quality: 202039 bases at least Q40		
	Consensus quality: 202804 bases at least Q30		
	Consensus quality: 203337 bases at least Q20		
	Estimated insert size: 175000; agarose-fp estimation		
	Estimated insert size: 206343; sum-of-contigs estimation		
	Quality coverage: 15.28 in Q20 bases; agarose-fp estimation		
	Quality coverage: 12.96 in Q20 bases; sum-of-contigs estimation.		
	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 7 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		
	1	1219: contig of 1219 bp in length	
*	1220	1319: gap of unknown length	
*	1320	2493: contig of 1174 bp in length	
*	2494	2593: gap of unknown length	
*	2594	3666: contig of 1073 bp in length	
*	3667	3766: gap of unknown length	


```

*      3767      5167: contig of 1401 bp in length
*      5168      5267: gap of unknown length
*      5268      12818: contig of 7551 bp in length
*      12918      36168: gap of unknown length
*      12919      36168: contig of 23251 bp in length
*      36170      36269: gap of unknown length
*      36270      206943: contig of 170674 bp in length.
FEATURES
    source
        1. 206943
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="RP11-1301B21"
            /clone_lib="RP11 human BAC library 11"
BASE COUNT   56312 a 44122 c 46163 g 59225 t 1121 others
ORIGIN
Query Match      78.2%; Score 1597; DB 2; Length 206943;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
*      107      AAGCTTCCACCTTACAGACAAATGCTACTGCTTAAAAAATGAAAGCTCTGCTGTTCTG 166
*      125368   AAGCTTCCACCTTACAGACAAATGCTACTGCTTAAAAAATGAAAGCTCTGCTGTTCTG 125309
*      167      GTTCCAGATGGCATCTGGCTCTATCTTCCACATGTACAGCCAAACATACAGCTCC 226
*      125308   GTTCCAGATGGCATCTGGCTCTATCTTCCACATGTACAGCCAAACATACAGCTCC 125249
*      227      CTGTCTATGAAGGACACAGCCGAGCATGACAGCTGCTGTCTGTTCTTCCGTGGCGCTCT 286
*      125248   CTGTCTATGAAGGACACAGCCGAGCATGACAGCTGCTGTCTGTTCTTCCGTGGCGCTCT 125189
*      287      GGCCTCTCTTTTGTGGGGACCTTTTGGGCGACACCAGATGTTTCTACCTGATGGAG 346
*      125188   GGCCTCTCTTTTGTGGGGACCTTTTGGGCGACACCAGATGTTTCTACCTGATGGAG 125129
*      347      CCGGCGGAGGACGCTGGATGACCTTGAAGAGACACCGGCTGGATGCTGCAATGGCT 406
*      125128   CCGGCGGAGGACGCTGGATGACCTTGAAGAGACACCGGCTGGATGCTGCAATGGCT 125069
*      407      GTGGGGATCTGATACGGGCGCTCTTGTGCGACATGAGACGCTTTGATGCTTACATG 466
*      125068   GTGGGGATCTGATACGGGCGCTCTTGTGCGACATGAGACGCTTTGATGCTTACATG 125009
*      467      GAACCTGTGCTCCCGGAGACAGTCAGCCTCTTTCAGTGGGAGAAACGCCGGCCCTGTGT 526
*      125008   GAACCTGTGCTCCCGGAGACAGTCAGCCTCTTTCAGTGGGAGAAACGCCGGCCCTGTGT 124949
*      527      TCTGCACCTGCTGTGACATCATCCCAAGATGAATCATCCCGGGGCTCACTGCAGG 586
*      124948   TCTGCACCTGCTGTGACATCATCCCAAGATGAATCATCCCGGGGCTCACTGCAGG 124889
*      587      CTCCTGTGACAGTCAACAGCCCTTGTGAGGTGTGAGAGAGCCCTCCCTCTACAGCCAC 646
*      124888   CTCCTGTGACAGTCAACAGCCCTTGTGAGGTGTGAGAGAGCCCTCCCTCTACAGCCAC 124829
*      647      GTGCTGCTCAAGAGAGTGCGCTTCTTCAACTGCAGTCCCTCTACCCGCTGCTGAAGAC 706
*      124828   GTGCTGCTCAAGAGAGTGCGCTTCTTCAACTGCAGTCCCTCTACCCGCTGCTGAAGAC 124769
*      707      CCCCTCCCTCAACGCAATGATGTGACAGCTGTCCGGGAAACCCCGGGGCGCTTCCGCTTCC 766
*      124768   CCCCTCCCTCAACGCAATGATGTGACAGCTGTCCGGGAAACCCCGGGGCGCTTCCGCTTCC 124709
*      767      CGAGAACGCAAAAGGAGATCTCATATTGACAGTGCATTTGATGGGGACAGATGAG 826
*      124708   CGAGAACGCAAAAGGAGATCTCATATTGACAGTGCATTTGATGGGGACAGATGAG 124649
*      827      CAAAACCTCAAGAAAGGAGCAACCTCTATATGTATGACAGGTGATCTGCAAAAGCCAG 886
*      124648   CAAAACCTCAAGAAAGGAGCAACCTCTATATGTATGACAGGTGATCTGCAAAAGCCAG 124589

```

```

OY      887      CTGAGATCTCAAGACATCCAGTCCCTGGCCCAAGGCCCTGCAGAACGCTACTGCTT 946
DB      124588   CTGAGATCTCAAGACATCCAGTCCCTGGCCCAAGGCCCTGCAGAACGCTACTGCTT 124529
OY      947      GTGGGCTATGAGAGCAGCTGGCTCGAGCCCTGTGGCCAGAGACTTCCGAAATGATGAATTC 1006
DB      124528   GTGGGCTATGAGAGCAGCTGGCTCGAGCCCTGTGGCCAGAGACTTCCGAAATGATGAATTC 124469
OY      1007     GTGGGATTTGGAATTTCTGGCCCATCTTCAAGACCTGGGTGCTATACATCACCAGGCAAG 1086
DB      124468   GTGGGATTTGGAATTTCTGGCCCATCTTCAAGACCTGGGTGCTATACATCACCAGGCAAG 124409
OY      1067     GGCATGGGTGACACAGGCTTTCCACACAAATCCAGAGGATGCGCTTAATGTTCCAGAGCT 1126
DB      124408   GGCATGGGTGACACAGGCTTTCCACACAAATCCAGAGGATGCGCTTAATGTTCCAGAGCT 124349
OY      1127     TGCGGCTGTGCTTGGCCCTATGAAAGGTTTCTGACTTCAGAAAGCCTGTGGCGATGCC 1186
DB      124348   TGCGGCTGTGCTTGGCCCTATGAAAGGTTTCTGACTTCAGAAAGCCTGTGGCGATGCC 124289
OY      1187     ATGAATTTGCTGGGCTACCGGACGTCAGATCTGAACAAGAACAGAAACCTTTGCTG 1246
DB      124288   ATGAATTTGCTGGGCTACCGGACGTCAGATCTGAACAAGAACAGAAACCTTTGCTG 124229
OY      1247     GATCTCTGTCTACTGCTGAGCTGTCCCTGAGCAATCCACTAAGAGGGTTGAAGGCTTT 1306
DB      124228   GATCTCTGTCTACTGCTGAGCTGTCCCTGAGCAATCCACTAAGAGGGTTGAAGGCTTT 124169
OY      1307     GCTGCCACCTGTGCTGACGCTCAGTCACCTTCTCTGATGCTTGTGAGCCTTGCTACAT 1366
DB      124168   GCTGCCACCTGTGCTGACGCTCAGTCACCTTCTCTGATGCTTGTGAGCCTTGCTACAT 124109
OY      1367     CTCTGAGCCTTAACCTACATGCTGTGGGTATCACACTGAGTGTGAGTTGTGCCACAGT 1426
DB      124108   CTCTGAGCCTTAACCTACATGCTGTGGGTATCACACTGAGTGTGAGTTGTGCCACAGT 124049
OY      1427     GCTCAAGAGAGAGACTTTTGTGTCATGCTTGTGTAGAAACAGACCTGGGAACTT 1486
DB      124048   GCTCAAGAGAGAGACTTTTGTGTCATGCTTGTGTAGAAACAGACCTGGGAACTT 123989
OY      1487     ATGTGACAGACATCCACCAGTAGAAACAGAGGATGCTCTCTTCTTCTGATCTT 1546
DB      123988   ATGTGACAGACATCCACCAGTAGAAACAGAGGATGCTCTCTCTTCTTCTGATCTT 123929
OY      1547     CCTGTCTGGGACACTTCAGAGACTTTGTGGCTGTGGAGGCTATTAAAGCAGACAGATA 1606
DB      123928   CCTGTCTGGGACACTTCAGAGACTTTGTGGCTGTGGAGGCTATTAAAGCAGACAGATA 123869
OY      1607     TCAGTGGAAATGATCCATTAACCTCCCGTGTCCACATCTTGCCCAATGGGAAATGATCTT 1666
DB      123868   TCAGTGGAAATGATCCATTAACCTCCCGTGTGTCCACATCTTGCCCAATGGGAAATGATCTT 123809
OY      1667     TCACCAAAAGACCTCACAGCATTTTCCACAGAGATGC 1703
DB      123808   TCACCAAAAGACCTCACAGCATTTTCCACAGAGATGC 123772

RESULT 6
LOCUS      BD127258      1979 bp      DNA      linear      PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION  BD127258
VERSION    BD127258.1 GI:33222203
KEYWORDS   JP 200201375-A/2689.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
1 (bases 1 to 1979)
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.

```


TITLE Primer for synthesizing full-length cDNA and use thereof
JOURNAL Patent: JP 2002017375-A 2689 22-JAN-2002;
COMMENT HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/2689
PD 22-JAN-2002
PF 07-JUL-2000 JP 2000253172
PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
PI ISHII,
PI YURI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI, HISASHI KOGA
PC
C12N15/09, C07K14/47, C07K16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/PC
10, C12P21/02, C12Q1/68//C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT CDS Location/Qualifiers
1.1979
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 451 a 542 c 474 g 512 t
ORIGIN
Query Match 75.7%; Score 1547; DB 6; Length 1979;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; gaps 0;
QY 106 CAAGGCTTCACCTAGACATGCTACGCTAAATAAATGAACCTCTGCTGTTCT 165
DB 47 CAAAGCTTCACCTAGACATGCTACGCTAAATAAATGAACCTCTGCTGTTCT 106
QY 166 GGTTCACAGATGGCCATCTTGCTATCTTCCACATGTACAGCCACAACTACAGTC 225
DB 107 GGTTCACAGATGGCCATCTTGCTATCTTCCACATGTACAGCCACAACTACAGTC 166
QY 226 CCGTGTATGAAAGGACACAGCCGAGCGCATGACGTCGTGTTCTCTTCTGCGGCTC 285
DB 167 CCGTGTATGAAAGGACACAGCCGAGCGCATGACGTCGTGTTCTCTTCTGCGGCTC 226
QY 286 TGCCCTCTCTTTGTGGGAGGCTTTTGGGACAGCCAGATGTTTCTACTGTAATGA 345
DB 227 TGCCCTCTCTTTGTGGGAGGCTTTTGGGACAGCCAGATGTTTCTACTGTAATGA 286
QY 346 GCCCGCTGGCAGTGTGATGACCTTCAAGCAGACACCGCTGATGCTGCACATGGC 405
DB 287 GCCCGCTGGCAGTGTGATGACCTTCAAGCAGACACCGCTGATGCTGCACATGGC 346
QY 406 TGTGCGGATCTGATACGGGCGCTTCTTGTGCAATGACGTCCTTGTGATGCTTACAT 465
DB 347 TGTGCGGATCTGATACGGGCGCTTCTTGTGCAATGACGTCCTTGTGATGCTTACAT 406
QY 466 GGAACCTGATCCCGGAGACAGTCAGCCTTTTCAAGTGGAGAAAGCCGGGCGCTG 525
DB 407 GGAACCTGATCCCGGAGACAGTCAGCCTTTTCAAGTGGAGAAAGCCGGGCGCTG 466
QY 526 TTCTGCACTGCTGTGACATCATCCACAAAGATGAATCATCCCGGGCTACTGCA 585
DB 467 TTCTGCACTGCTGTGACATCATCCACAAAGATGAATCATCCCGGGCTACTGCA 526
QY 586 GCTCCTGTGACAGCAAGCCCTTTGAGGTGTGAGAAAGCCCTGCGCTCTACAGCA 645
DB 527 GCTCCTGTGACAGCAAGCCCTTTGAGGTGTGAGAAAGCCCTGCGCTCTACAGCA 586
QY 646 CGTGCTGCTCAAGAGAGTGCCTTCTTCAACTGCACTGCTTACCGCTGCTGAAGA 705
DB 587 CGTGCTGCTCAAGAGAGTGCCTTCTTCAACTGCACTGCTTACCGCTGCTGAAGA 646
QY 706 CCCCTCCCAACCTGATATGCTGCACTGCTGCGGAGCCCGGGCGGTGCTCCGTT 765

DB 647 CCCCTCCCAACCTGATATGCTGCACTGCTGCGGAGCCCGGGCGGTGCTCCGTT 706
QY 766 CCGAGAACGCACAAAGAGATCTCATGATTGACATCCCATTTGATGGGAGCATGA 825
DB 707 CCGAGAACGCACAAAGAGATCTCATGATTGACATCCCATTTGATGGGAGCATGA 766
QY 826 GCAAAAACCTCAGAGAGAGACCAACCTACTATGTGATGACAGTCATCTGCCAAGCA 885
DB 767 GCAAAAACCTCAGAGAGAGACCAACCTACTATGTGATGACAGTCATCTGCCAAGCA 826
QY 886 GCTGAGATCTCAAGACCATCCAGTCCTTGGCCAGAGCCCTGACAGACGCTACTGCT 945
DB 827 GCTGAGATCTCAAGACCATCCAGTCCTTGGCCAGAGCCCTGACAGACGCTACTGCT 886
QY 946 TGTGCGTATGAGAGACCTGCTGAGCCCTGCTGCGCCAGACTTCCGATGTATGAAT 1005
DB 887 TGTGCGTATGAGAGACCTGCTGAGCCCTGCTGCGCCAGACTTCCGATGTATGAAT 946
QY 1006 CGTGGATTTGGAATTTCTGCCCCATCTTCAAGACCTGGGTGATACATCACCGAGGCA 1065
DB 947 CGTGGATTTGGAATTTCTGCCCCATCTTCAAGACCTGGGTGATACATCACCGAGGCA 1006
QY 1066 GGGCATGGGTGACACGCTTTCACACAAATGCGAGGAGCCCTTAATGTCTCCAGGC 1125
DB 1007 GGGCATGGGTGACACGCTTTCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGC 1066
QY 1126 TTGGCGCTGCTTTGCCCTATGAAAAGTTTCTGACTTTCAGAAAGCCTGTGCGATGC 1185
DB 1067 TTGGCGCTGCTTTGCCCTATGAAAAGTTTCTGACTTTCAGAAAGCCTGTGCGATGC 1126
QY 1186 CATGAATTTGTGGGTACCGCCAGCTCGATGTGAACAAGAACAGAAACCTGTGCT 1245
DB 1127 CATGAATTTGTGGGTACCGCCAGCTCGATGTGAACAAGAACAGAAACCTGTGCT 1186
QY 1246 GGATCTTCTGTACCTGAGCTGTCCCTGAGCAAACTCACTAAGAGGTTGAGAAAGCTT 1305
DB 1187 GGATCTTCTGTACCTGAGCTGTCCCTGAGCAAACTCACTAAGAGGTTGAGAAAGCTT 1246
QY 1306 TGTGCGACCTGCTGACGCTCAGTACCTTCTGTAATGCTTGTGAGCCTTGCTTACA 1365
DB 1247 TGTGCGACCTGCTGACGCTCAGTACCTTCTGTAATGCTTGTGAGCCTTGCTTACA 1306
QY 1366 TGTGCGACCTTACACATGCTGTGGGATACACTGAGTGTGATGCTGCTGACAG 1425
DB 1307 TGTGCGACCTTACACATGCTGTGGGATACACTGAGTGTGATGCTGCTGACAG 1366
QY 1426 TGTCAAGCAGAAAGACTTTTGTTCATGCTTGTGTCTAGAAAACAGACTGGGAACT 1485
DB 1367 TGTCAAGCAGAAAGACTTTTGTTCATGCTTGTGTCTAGAAAACAGACTGGGAACT 1426
QY 1486 TATGTAGCAGCAGATCCACAGTGAAGAGGATATGCTCTTCTTCTTCTTCTGATCT 1545
DB 1427 TATGTAGCAGCAGATCCACAGTGAAGAGGATATGCTCTTCTTCTTCTTCTGATCT 1486
QY 1546 TCTGTCTGGGAGACTTACAGAGACTTGTGGCTGGAGGCCATTAACAGACAGCAT 1605
DB 1487 TCTGTCTGGGAGACTTACAGAGACTTGTGGCTGGAGGCCATTAACAGACAGCAT 1546
QY 1606 ATCAGTGAATGATCATTAACCTCCCTGTCACACTTGTGCCAATGGGGAATGATCT 1665
DB 1547 ATCAGTGAATGATCATTAACCTCCCTGTCACACTTGTGCCAATGGGGAATGATCT 1606
QY 1666 TTCACCAAGAGCTACACAGCATTTTTCACAGAGATGC 1703
DB 1607 TTCACCAAGAGCTACACAGCATTTTTCACAGAGATGC 1644
RESULT 7
AK074746 1979 bp mRNA linear PRI 03-SEP-2002
LOCUS AK074746
DEFINITION Homo sapiens cDNA FLJ90265 f1s, clone NTBRM4001325, weakly similar
TO CHONDROITIN 6-SULFOTRANSFERASE (EC 2.8.2.17).
ACCESSION AK074746

VERSION AK074746.1 GI:22760388
 KEYWORDS oligo capping; f1s (full insert sequence).
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 REFERENCE 1
 AUTHORS Iisogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T.,
 Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
 Kawai,Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
 Kojima,S., Nagahara,K., Masuno,Y., Ono,T., Okano,K., Yoshikawa,Y.,
 Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
 Ninomiya,K.
 TITLE NEDO human cDNA sequencing project
 JOURNAL unpublished
 REFERENCE 2 (bases 1 to 1979)
 AUTHORS Iisogai,T. and Otsuki,T.
 JOURNAL Direct Submission
 Submitted (25-MAR-2002) Takao Iisogai, Helix Research Institute,
 Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 297-0812, Japan
 (E-mail:genomics@hi.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
 COMMENT NEDO human cDNA sequencing project supported by Ministry of
 Economy, Trade and Industry of Japan; cDNA full insert sequencing:
 Research Association for Biotechnology; cDNA library construction:
 Institute of Medical Science, University of Tokyo, Laboratory of
 Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
 sequencing and clone selection: Helix Research Institute (supported
 by Japan Key Technology Center etc.).
 FEATURES
 source Location/Qualifiers
 1..1979
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="NT2RM4001325"
 /cell_line="NT2"
 /cell_type="teratocarcinoma"
 /clone_id="NT2RM4"
 /note="cloning vector: PME18SFL3-mRNA from uninduced NT2
 neuronal precursor cells"
 69..1229
 /note="unamed protein product"
 /codon_start=1
 /protein_id="BAC11177.1"
 /db_xref="GI:22760389"
 /translation="MLPGRKKLLFLYSOMAILALFPHMYSHNISLSMAKOPRHH
 VIVTSSMRSGSFYGLGHPDVFYLMPEPMHMYMPEKFSFAMLLMAYVDLIRAYF
 LCDMSYVDATHEPRPROSSLFQENSRALCSAPACITIPDEITIPRAKCLTSQDP
 FEVEKACRSTSHVVLKVFREFNLQSLPLKDPDLNHLIVLRYDPAYFRSRERK
 GDLMDIRIVWGQHEOKLKEEDOPYVMQVQSQLELYTKTIOSLPRALORLYLVRY
 EDLARAVVAQTSRMVEFVLEFLPLQTVHNITRGKMGMDHAFHTVARDLNYSQAW
 RMSLPYKVSLSLQKACGDAMNLLGVRHVRSPQEDRNILLDELSTWVPEQTH"
 CDS
 BASE COUNT 451 a 542 c 474 g 512 t
 ORIGIN
 Query Match 75.7%; Score 1547; DB 9; Length 1979;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 346 GCCGCCCTGGCAGCTGTGTGATGACCTTCAGCAGACGCGCTGGATGCTGCACATGGC 405
 DB 287 GCGCGCTGGCAGCTGTGTGATGACCTTCAGCAGACGCGCTGGATGCTGCACATGGC 346
 QY 406 TGTGCGGGATGTGATACGGCGCTCTTGTGCGACATGAGCGCTTTGATGGCTACAT 465
 DB 347 TGTGCGGGATGTGATACGGCGCTCTTGTGCGACATGAGCGCTTTGATGGCTACAT 406
 QY 466 GGAACCTGTGTCGGGACAGACAGTCACAGCTCTTTTCAGTGGAGAAACACCGGGCCCTGTG 525
 DB 407 GGAACCTGTGTCGGGACAGACAGTCACAGCTCTTTTCAGTGGAGAAACACCGGGCCCTGTG 466
 QY 526 TTTCGACACCTGCTGTGATCATCATCCCAAGATGAATCATCCCGGGGCTGACTGAG 585
 DB 467 TTTCGACACCTGCTGTGATCATCATCCCAAGATGAATCATCCCGGGGCTGACTGAG 526
 QY 586 GCTCCTGTGACATCAACAGCCCTTTGAGTGGTGGAGAAGGCTGCGCTCTACAGCA 645
 DB 527 GCTCCTGTGACATCAACAGCCCTTTGAGTGGTGGAGAAGGCTGCGCTCTACAGCA 586
 QY 646 CGTGTGCTCAAGAGAGTGCGCTTCTTCAACCTGCAGTCCCTTACCCGCTGTGAAGA 705
 DB 587 CGTGTGCTCAAGAGAGTGCGCTTCTTCAACCTGCAGTCCCTTACCCGCTGTGAAGA 646
 QY 706 CCCCTCCCTCAACCTGCATATCGTGCACCTGTCGGGAGCCCGGGGCGCTTCCGTTG 765
 DB 647 CCCCTCCCTCAACCTGCATATCGTGCACCTGTCGGGAGCCCGGGGCGCTTCCGTTG 706
 QY 766 CCGAGAACGCAAAAGGAGATCTCATGATTTGACAGTGCATTTGATGGGCGAGCATGA 825
 DB 707 CCGAGAACGCAAAAGGAGATCTCATGATTTGACAGTGCATTTGATGGGCGAGCATGA 766
 QY 826 GCAAAACTCAAGAAAGGAGACCAACCTACTATGTATGTGACAGTGCATCTGCCAAGCA 885
 DB 767 GCAAAACTCAAGAAAGGAGACCAACCTACTATGTATGTGACAGTGCATCTGCCAAGCA 826
 QY 886 GCTGAGATCTACAAAGACATCATCAGTCTTCCCAAGGCGCTGAGAAACCTTACCTCT 945
 DB 827 GCTGAGATCTACAAAGACATCATCAGTCTTCCCAAGGCGCTGAGAAACCTTACCTCT 886
 QY 946 TGTGCGCTATGAGACCTGCTGAGCGCCCTGTGGCCGACAGTCCGGAATGTATGAAT 1005
 DB 887 TGTGCGCTATGAGACCTGCTGAGCGCCCTGTGGCCGACAGTCCGGAATGTATGAAT 946
 QY 1006 CGTGGATGTGAATTTCTGCCCCATCTTCAACACCTGGGTGGATTAACATCACCCAGCAA 1065
 DB 947 CGTGGATGTGAATTTCTGCCCCATCTTCAACACCTGGGTGGATTAACATCACCCAGCAA 1006
 QY 1066 GGGCATGGTACCAACAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGC 1125
 DB 1007 GGGCATGGTACCAACAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGC 1066
 QY 1126 TTGGCGCTGTGTTTCCCTATGAAAGGTTTCTGACTGAGAAAGCCTGTGGCGATGC 1185
 DB 1067 TTGGCGCTGTGTTTCCCTATGAAAGGTTTCTGACTGAGAAAGCCTGTGGCGATGC 1126
 QY 1186 CATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAAGAACAAGAAACCTGTGCT 1245
 DB 1127 CATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAAGAACAAGAAACCTGTGCT 1186
 QY 1246 GGATCTTCTGTACCTGAGATGTCCTGAGCAATATCCATAGAGGGTTGAGAAGGCTT 1305
 DB 1187 GGATCTTCTGTACCTGAGATGTCCTGAGCAATATCCATAGAGGGTTGAGAAGGCTT 1246
 QY 1306 TGTGTCACCTGTGTGACGCTCAAGTCACTTTCTGTGATGCTTGTGAGCTTGTGCTACA 1365
 DB 1247 TGTGTCACCTGTGTGACGCTCAAGTCACTTTCTGTGATGCTTGTGAGCTTGTGCTACA 1306
 QY 1366 TCTCTGACCTTAATCAATGCTGTGGGTATCAACAGATGAGATGTGTGCTACACAG 1425
 DB 1307 TCTCTGACCTTAATCAATGCTGTGGGTATCAACAGATGAGATGTGTGCTACACAG 1366

OY	1426	TGCTCAAGCAGAAGAGACTTTTGTGTCCATGGTTTGTCCTAAGAAAACAGACTGGGGAACT	1485
Dd	1367	TGCCTAAGCAGAAGAGACTTTTGTGTCCATGGTTTGTCCTAAGAAAACAGACTGGGGAACT	1426
OY	1486	TATGTAGACGACACATCCCACCAGTAGAACAAGSGATTGCTCTTCTTTCTTGATCT	1545
Dd	1427	TATGTAGACGACACATCCCACCAGTAGAACAAGSGATTGCTCTTCTTTCTTGATCT	1486
OY	1546	TCCGTGTGGGCGACACTTCAGAGACTTTGTGGCGCTGGAGGCGCTATTGAAGCACGACAGT	1605
Dd	1487	TCCGTGTGGGCGACACTTCAGAGACTTTGTGGCGCTGGAGGCGCTATTGAAGCACGACAGT	1546
OY	1606	ATCAGTGGAAATTGATTCATAAAGCTCCCTGTCCACATCTTGCCCAATGGGGAATGATCT	1665
Dd	1547	ATCAGTGGAAATTGATTCATAAAGCTCCCTGTCCACATCTTGCCCAATGGGGAATGATCT	1606
OY	1666	TTTACCAAAGAGCTCACACGCAATTTCCACAGAGATGC	1703
Dd	1607	TTTACCAAAGAGCTCACACGCAATTTTCCACAGAGATGC	1644

[illegible]

OY	526	TTCTGACACCTGCTGTGACATCATCCCAAGATATAAATCATCCCGGGCTCACTGGAC	585
Db	580	TTCTGACACCTGCTGTGACATCATCCCAAGATATAAATCATCCCGGGCTCACTGGAC	639
OY	586	GCTCTGTGTGACATCAACAGCCCTTTAGGTGGTGAGAGAGGCCCTGCTCTCTACAGCCA	645
Db	640	GCTCTGTGTGACATCAACAGCCCTTTAGGTGGTGAGAGAGGCCCTGCTCTCTACAGCCA	699
OY	646	GCTGGTGTCAAGGAGGGGCGCTTTTCAACCTGAGAGCCCTCATCCCGCTGTAAAGA	705
Db	700	GCTGGTGTCAAGGAGGGGCGCTTTTCAACCTGAGAGCCCTCATCCCGCTGTAAAGA	759
OY	706	CCCTCCCTCAACCTGCATATCTGTGCACCTGGTCCGGAGACCCCGGGCGTGTCCGTTT	765
Db	760	CCCTCCCTCAACCTGCATATCTGTGCACCTGGTCCGGAGACCCCGGGCGTGTCCGTTT	819
OY	766	CCGAGAACGCACAAAGGAGAGATCTCATATTGACAGTGGCATTTGTGATGGGGCACACATGA	825
Db	820	CCGAGAACGCACAAAGGAGAGATCTCATATTGACAGTGGCATTTGTGATGGGGCACACATGA	879
OY	826	GCAAAAACTCAAGAAAGAGAGACAACCCCTACTATGTGATGAGAGTCATCTGCCAAAGCCA	885
Db	880	GCAGAACTCAAGAAAGAGAGACCAACCCCTACTATGTGATGAGAGTCATCTGCCAAAGCCA	939
OY	886	GCTGAGATCTTACAAAGACCATCCAGTCTTGGCCAAAGCCCTGCAGAGAACGCTACTCTGT	945
Db	940	GCTGAGATCTTACAAAGACCATCCAGTCTTGGCCAAAGCCCTGCAGAGAACGCTACTCTGT	999
OY	946	TGTGGGCTATGAGGACCTGGGCTCGAGCCCTGTGGCCCGACAGACTTCCGAAATGTAATTT	1005
Db	1000	TGTGGGCTATGAGGACCTGGGCTCGAGCCCTGTGGCCCGACAGACTTCCGAAATGTAATTT	1055
OY	1006	CGTGGGATTTGGAATTTCTTGCCCATCTTCAAGACCTGGGTGCATTAACATCACCCAGGCCA	1065
Db	1060	CGTGGGATTTGGAATTTCTTGCCCATCTTCAAGACCTGGGTGCATTAACATCACCCAGGCCA	1115
OY	1066	GGGATATGGGTGACACACGCTTTCCACACAAATGCCAGGGATGCCCTTAATGTCTCCACAGC	1122
Db	1120	GGGATATGGGTGACACACGCTTTCCACACAAATGCCAGGGATGCCCTTAATGTCTCCACAGC	1179
OY	1126	TTGGGCGCTGTCTTGGCCCTATGAAAAGAGTTTCTGACCTTCAGAAAGCCTGGGGGATGC	1185
Db	1180	TTGGGCGCTGTCTTGGCCCTATGAAAAGAGTTTCTGACCTTCAGAAAGCCTGGGGGATGC	1233
OY	1186	CATGAATTTGCTGGGGCTACGCCACGTCAGATCTGAAACAGAAACAGAAACCTGTGCT	1245
Db	1240	CATGAATTTGCTGGGGCTACGCCACGTCAGATCTGAAACAGAAACAGAAACCTGTGCT	1299
OY	1246	GGATCTTGTGTCTAACCCTGGACTGTCCCTGACACAAATCCACTTAAGAGGGTTAGAAAGGCTT	1305
Db	1300	GGATCTTGTGTCTAACCCTGGACTGTCCCTGACACAAATCCACTTAAGAGGGTTAGAAAGGCTT	1355
OY	1306	TGCTGCCACCGTGTGTGACCTCAGTCACCTTCTCTGTGAATGCTTGTGACCTTGGCTGTACA	1365
Db	1360	TGCTGCCACCGTGTGTGACCTCAGTCACCTTCTCTGTGAATGCTTGTGACCTTGGCTGTACA	1419
OY	1366	TCTCTGAGCCTTAACTACATATCTGTGTGGGTATCACACTGAGTGTGATGTGTCCACAGC	1422
Db	1420	TCTCTGAGCCTTAACTACATATCTGTGTGGGTATCACACTGAGTGTGATGTGTGTCCACAGC	1479
OY	1426	TGCTCAAGCACAAGAGACTTTGTGTGCATGCTGTGTGTAGAAAAACAGCTGGGGAGACT	1485
Db	1480	TGCTCAAGCACAAGAGACTTTGTGTGCATGCTGTGTGTGTAGAAAAACAGCTGGGGAGACT	1533
OY	1486	TATGTGAGACAGACATCCACAGTGAAGAGGGATTTGCTCTTCTTCTTGTGATCT	1545
Db	1540	TATGTGAGACAGACATCCACAGTGAAGAGGGATTTGCTCTTCTTCTTGTGATCT	1599
OY	1546	TCTCTGTCTGGGACACTTCAAGACTTTGTGGCTCGAGAGGCTTATTAAGCAGACACAGT	1605
Db	1600	TCTCTGTCTGGGACACTTCAAGACTTTGTGGCTCGAGAGGCTTATTAAGCAGACACAGT	1659

```

QY 1606 ATCAAGTGAATTCATCAATAACCTCCTGTCCATCTGCCAATGGGAATGATCT 1665
DB 1660 ATCAAGTGAATTCATCAATAACCTCCTGTCCATCTGCCAATGGGAATGATCT 1719
QY 1666 TTCACCAAGAGCTTACCAGCATTTTCCACAGAGATGC 1703
DB 1720 TTCACCAAGAGCTTACCAGCATTTTCCACAGAGATGC 1757

RESULT 9
BD134772 2032 bp DNA linear PAT 18-SEP-2002
LOCUS Glycosylsulfoltransferase-3.
DEFINITION BD134772
ACCESSION BD134772.1 GI:32229717
VERSION JP 2002507409-A/1.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistrup, A., Rosen, S.D., Tangemann, K. and Hemmerich, S.
TITLE Glycosylsulfoltransferase-3
JOURNAL Patent: JP 2002507409-A 1 12-MAR-2002;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, SYNTAX INC
COMMENT OS Homo sapiens (human)
PN JP 2002507409-A/1
PD 12-MAR-2002
PR 26-FEB-1999 JP 2000537979
PR 20-MAR-1998 US 09/045284,12-NOV-1998 US 09/190911 PI
ANNETTE BISTRUP, STEVEN D ROSEN, KIRSTEN TANGEMANN, STEFAN PI
HEMMERICH
PC C12N1/10, A01K67/027, A61K45/00, A61K48/00, A61P29/00, A61P37/06,
PC A61P43/00,
PC C12N1/15, C12N1/21, C12N5/10, C12N15/09, C12Q1/48, C12N5/00, C12N15/
PC 00
CC Glycosylsulfoltransferase-3
C1 key Location/Qualifiers
FT source 1.2032
FT location/Qualifiers
1.2032 location/Qualifiers
/organism="Homo sapiens (human)"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN
Query Match 75.7%; Score 1547; DB 6; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```

```

DB 460 TGTGGGATCTGATACAGGGCGCTTCTGTGGACATGAGCGCTTGTGATGCTTACAT 519
QY 466 GGAACCTGGTCCCGGGAGACAGTCCAGCCTCTTTCAGTGGGAGAAACAGCGGGCCCTGTG 525
DB 520 GGAACCTGGTCCCGGGAGACAGTCCAGCCTCTTTCAGTGGGAGAAACAGCGGGCCCTGTG 579
QY 526 TTTGACACCTGCTGATGATCATCATCCCAAGATGAATCCCGGGGCTGACATGAG 585
DB 580 TTTGACACCTGCTGATGATCATCATCCCAAGATGAATCCCGGGGCTGACATGAG 639
QY 586 GCTCCTGTGACATCAAGACCCCTTTCAGTGGTGGAGAGAGCCCTCCCTCTACAGCA 645
DB 640 GCTCCTGTGACATCAAGACCCCTTTCAGTGGTGGAGAGAGCCCTCCCTCTACAGCA 699
QY 646 GCTGCTGTCAAGAGAGTGGCTTTTCAACCTGCAATGCTTACCGGCTGCTAAGA 705
DB 700 GCTGCTGTCAAGAGAGTGGCTTTTCAACCTGCAATGCTTACCGGCTGCTAAGA 759
QY 706 CCCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGGCGGTTCGTTTC 765
DB 760 CCCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGGCGGTTCGTTTC 819
QY 766 CCGAGAACGCAAAAGGAGATCTCATGATGACAGTGCATTTGATGGGCGACATGA 825
DB 820 CCGAGAACGCAAAAGGAGATCTCATGATGACAGTGCATTTGATGGGCGACATGA 879
QY 826 GCAAAATCAAGAGAGAGAGACCAACCTTACTATGTATGCAAGCTATCTGCCAAGCA 885
DB 880 GCAAAATCAAGAGAGAGAGACCAACCTTACTATGTATGCAAGCTATCTGCCAAGCA 939
QY 886 GCTGAGATCTCAAGAGACATCACTGCTTCCAGGAGCCCGAGAAACGCTACTCTGT 945
DB 940 GCTGAGATCTCAAGAGACATCACTGCTTCCAGGAGCCCGAGAAACGCTACTCTGT 999
QY 946 TGTGGCTATGAGAGACCTGGCTGAGCCCTGTGCCCCAGACTTCCGAAATGATGAAT 1005
DB 1000 TGTGGCTATGAGAGACCTGGCTGAGCCCTGTGCCCCAGACTTCCGAAATGATGAAT 1059
QY 1006 GCTGGGATTTGGAAATCTTGGCCCATCTTCAAGACCTGGGAGCAATACATCAACCGAGCA 1065
DB 1060 GCTGGGATTTGGAAATCTTGGCCCATCTTCAAGACCTGGGAGCAATACATCAACCGAGCA 1119
QY 1066 GGGCATGGGTGACACAGCTTTCACACAAGATCCAGAGATGCTTAAATGTCTCCAGGC 1125
DB 1120 GGGCATGGGTGACACAGCTTTCACACAAGATCCAGAGATGCTTAAATGTCTCCAGGC 1179
QY 1126 TTTGGCTGTGCTTGTCCCTATGAAGATTTCTGACTTCAGAAAGCCTGTGGCATGC 1185
DB 1180 TTTGGCTGTGCTTGTCCCTATGAAGATTTCTGACTTCAGAAAGCCTGTGGCATGC 1239
QY 1186 CATGAATTTGGGGCTACCGGCACAGTCAAGATCTGAACAAGAAAGAGAAACCTGTGCT 1245
DB 1240 CATGAATTTGGGGCTACCGGCACAGTCAAGATCTGAACAAGAAAGAGAAACCTGTGCT 1299
QY 1246 GGAATCTTGTCTACCTGAGACTGCTCTGAGCAATTCACCTAAGAGGTTGAGAGGCTT 1305
DB 1300 GGAATCTTGTCTACCTGAGACTGCTCTGAGCAATTCACCTAAGAGGTTGAGAGGCTT 1359
QY 1306 TGTCTGACACTGTGTGACAGCTCACTTCTGTAATGCTTTCAGACCTTGTCTTACA 1365
DB 1360 TGTCTGACACTGTGTGACAGCTCACTTCTGTAATGCTTTCAGACCTTGTCTTACA 1419
QY 1366 TGTCTGACACTTAACTAATGCTGTGGGATCAACAGAGAGTGTGATGTTGTCACAG 1425
DB 1420 TGTCTGACACTTAACTAATGCTGTGGGATCAACAGAGAGTGTGATGTTGTCACAG 1479
QY 1426 TGTCTAAGCAGAGAGACTTTTGTGCATGCTTGTGTAGAAAACAGACTGGGAACT 1485
DB 1480 TGTCTAAGCAGAGAGACTTTTGTGCATGCTTGTGTAGAAAACAGACTGGGAACT 1539
QY 1486 TATGTAGACAGACATCCCAACAGTGAAGAAAGAGTATGCTTCTTTTGTGATCT 1545

```

Db	1540	AATGAGACAGACAATCCACCAGTGAANAACGGGATATGCCTTCCTCTTTCTTGTAAGCT	1599
Qy	1546	TCTGTCTGGGACAGACTTCAGAGACTTTGTGGCCTCGAGGCCATTAAAGCACAGACAGT	1605
Db	1600	TCTGTCTGGGACAGACTTCAGAGACTTTGTGGCCTCGAGGCCATTAAAGCACAGACAGT	1659
Qy	1606	ATCATGTGAATGATTCATAAAGCCCCCTGTCACATCTTCCCAATGGGAATGAGACT	1665
Db	1660	ATCATGTGAATGATTCATAAAGCCCCCTGTCACATCTTCCCAATGGGAATGAGACT	1719
Qy	1666	TTCACCAAAGAGCTCACCAGCAATTTCCACAGAGATGC	1703
Db	1720	TTCACCAAAGAGCTCACCAGCAATTTCCACAGAGATGC	1757
RESULT 10			
AFI31235		2032 bp mRNA linear PRI 10-AUG-1999	
FOCUS			
DEFINITION	Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA,		
complete cds.			
AFI31235			
AFI31235.1	GI:4927113		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
TITLE	1 (bases 1 to 2032) Bistrup,A., Bhakta,S., Lee,J.K., Belov,Y.Y., Gunn,M.D., Zuo,F.R., Huang,C.C., Kannagi,R., Rosen,S.D. and Hemmerich,S. Sulfotransferases of two specificities function in the reconstitution of high endothelial cell ligands for L-selectin J. Cell Biol. 145 (4), 899-910 (1999)		
JOURNAL	9926436		
MEDLINE	10330415		
PUBMED	2 (bases 1 to 2032)		
REFERENCE	Bistrup,A., Tangemann,K., Bhakta,S., Lee,J.-K., Belov,Y.Y., Gunn,M.D., Zuo,F.-R., Huang,C.-C., Kannagi,R., Rosen,S.D. and Hemmerich,S. Direct Submission Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience, 3401 Hillview Avenue, Palo Alto, CA 94304, USA		
AUTHORS	Location/Qualifiers		
FEATURES	1..2032		
source	/organism="Homo sapiens"		
	/mol_type="mRNA"		
	/db_xref="taxon:9606"		
	/cell_type="high endothelial"		
	/tissue_type="tonsil"		
	182..1342		
	/function="biosynthesis of endothelial ligands for		
	L-selectin		
	/note="HEC-GlcNAc-6-ST"		
	/codon_start=1		
	/product="N-acetylglucosamine 6-O-sulfotransferase"		
	/protein_id="AAD3015.1"		
	/db_xref="GI:4927114"		
	/translation="MLPKRKRLLLVLSOMALLFEFHMYSNISSLSKAQPERMH VLVLSWRSGSSEFGQLGFQGVDFYELMEPAHWMTFKOSTAMLMVAVDLRAVEFL LCDSVFDAVMPEGRPSILFQWENSRALCSAPACIIIPODETIIPAHOILCSODP FEVERACRSYSHYLKEVREPNOSIXPLKDPSLNHLHYLRDPRAYFRSERERK GDMLIDSRIVMGQHEOKLKEDDPITYMOWICQSOLITVKTQSLPALDERLYLVRY EDLARAVADTSRMTEFVGLEIFLEPHLOTWHNTIRKGMDGHAFHTNARALNVSQAM RMSLPYEKVSTSLAACCDAMNLGLGYRHVRSEOEORNLDDLSTWTVEQH"		
CDS			
	BASE COUNT	468 a 569 c 490 g 505 t	
ORIGIN			
Query Match	75.7%	Score 1547:	DB 9: Length 2032:
Best Local Similarity	99.9%:	Pred. No. 0:	
Matches 1597:	Conservative 0:	Mismatches 1:	Indels 0: Gaps 0:
QY	106	CAAGGTTCACATTACGACACATGCTACTGCTTAAAAAATGAAGTCCCTGCTTCTT	165

Db	160	C AAGTCTTCCACTTCACGACAAATGCTACTGCTTAAAAAAATGAAGTCTTCGTCTTCT	219
Qy	166	G G T T T C C A G A T G G C C A T C T T G G C C T A T T C T T Y C A C A T G T A C A C C A C A C A T C A G C T C	225
Db	220	G G T T T C C A G A T G G C C A T C T T G C T C T A T T C T T C C A C A T G T A C A C C A C A C A T C A G C T C	279
Qy	226	C C T G C T A T G A A G G C A C A G C C C G A C G C A T G C A C T G C T G C T C T G T C T C T C T G G C G C T C	285
Db	280	C C T G C T A T G A A G G C A C A G C C C G A C G C A T G C A C G T G C T C T G T C T C T C T G C G C T C	339
Qy	286	T G C T C T T C T T T T G T G G G G C A C C T T T T T G G G A G A C C A G A T G T T T T C T A C T G A T G A	345
Db	340	T G C T C T T C T T T T G T G G G G C A C C T T T T T G G G A G A C C A G A T G T T T C T A C T A T G A	399
Qy	346	G C C C G C C T T G G C A C G T G G A T G A C C T T C A A G C A G A C A C C G C C T G A T G C T G C A C A T G C	405
Db	400	G C C C G C C T T G G C A C G T G G A T G A C C T T C A A G C A G A C A C C G C C T G A T G C T G C A C A T G C	459
Qy	406	T G T G G G G A T C T G A T A C G G G C G T C T T C T T G T G C A C A T G A C G T C T T T G A T G C T A C A T	465
Db	460	T G T G G G G A T C T G A T A C G G G C G T C T T C T T G T G C A C A T G A C G T C T T T G A T G C T A C A T	519
Qy	466	G S A A C C T G T C C C C G G A G A C A G T C A G C C T C T T A G A G G G A A C A G C G G G C C C T G T G	525
Db	520	G S A A C C T G T C C C C G G A G A C A G T C A G C C T C T T A G A G G G A A C A G C G G G C C C T G T G	579
Qy	536	T T C T G C A C C T G C C T G A C A T C A T C C C A C A A G A T A A A T C A T C C C C G G G C T C A C T G C A G	585
Db	580	T T C T G C A C C T G C C T G A C A T C A T C C C A C A A G A T A A A T C A T C C C C G G G C T C A C T G C A G	639
Qy	586	G C T C C T G T G C A G T C A A C A G C C T T T G A G T G T G A A A G G C C T G C C G C T C T A C A G C C A	645
Db	640	G C T C C T G T G C A G T C A A C A G C C C T T T G A G T G T G A A A G G C C T C C C C T C T A C A G C C A	699
Qy	646	C G T G T G C T C A A G A G A G T G C G C T T T T A A C C T G T A G T C C C T C A C C G C T C T A A A A G A	705
Db	700	C G T G T G C T C A A G A G A G T G C G C T T T T A A C C T G T A G T C C C T C A C C G C T C T A A A A G A	759
Qy	706	C C C C T C C C T C A A C T G C A T A T C G T A C G C A C C T G T C C G G A C C C C G G G C G T T T C C G T T C	765
Db	760	C C C C T C C C T C A A C T G C A T A T C G T A C G C A C C T G T C C G G A C C C C G G G C G T T T C C G T T C	819
Qy	766	C C G A A A A G C A A A A G G A G A T C T A T A T T G A C A G T G C A T T G T G A T G G G G C A C A T A G A	825
Db	820	C C G A A A A G C A A A A G G A G A T C T A T A T T A G C A G T G C A T T G A T G G G G C A C A T A G A	879
Qy	826	G C A A A A A C T C A A G A G A G A C A C A C C T A C T A T G A T G A G A G T C A T C T G C A A A G C A	885
Db	880	G C A A A A A C T C A A G A G A G A C A C A C C T A C T A T G A T G C A G T C A T C T G C A A A G C A	939
Qy	886	G C T G A G A T C A C A A G A C A T C A G T C C T T G C C A A G G C C C T G C A G A A C G C T A C T C T	945
Db	940	G C T G A G A T C A C A A G A C A T C A G T C C T T G C C A A G G C C C T G C A G A A C G C T A C T C T	999
Qy	946	T G T G G C T A T A G A C C T G G C T C A G A G C C C T G T G G C C C A G A C T T C C C A A T G T A T G A A T T	1005
Db	1000	T G T G G C T A T A G A C C T G G C T C A G A G C C C C T G T G C C C A G A C T T C C A A A T G T A T G A A T T	1059
Qy	1006	C G T G G A T T G A A A T C T T G C C C A C T T C A G A C C T G G G T G A T A C A T C A C C C G A G G A A	1065
Db	1060	C G T G G A T T G A A A T C T T G C C C A C T T C A G A C C T G G G T G C A T A C A T C A C C C G A G G A A	1119
Qy	1066	G G G C A T G G G T G A C A C G C T T T C A C A C A A A T G C C A G G A T G C C C T T A A T G T C T C C A G C	1125
Db	1120	G G G C A T G G G T G A C A C G C T T T C A C A C A A A T G C C A G G A T G C C C T T A A T G T C T C C A G C	1179
Qy	1126	T T G G C G T G T C T T T G C C C T A T G A A A G T T T C T G A C T T A G A A A G C C T G T G C G A T G C	1185
Db	1180	T T G G C G T G T C T T T G C C C T A T G A A A G T T T C T G A C T T A G A A A G C C T G T G G G A T G C	1239
Qy	1186	C A T G A A T T G C T G G G C T A C C G C A C A G T C A G A T C T G A A A A A C A A C A A A A C T G T T G C T	1245
Db	1240	C A T G A A T T G C T G G G C T A C C G C A C A G T C A G A T C T G A A A A A A A C A A A A C A A A C T G T T G C T	1299

QY 1246 GGATCTTCTGTCTACTGAGCTGTCCTGAGCAATCCACTAGAGGTTGAGAAGCTT 1305
 |||||||
 Db 1300 GGATCTTCTGTCTACTGAGCTGTCCTGAGCAATCCACTAGAGGTTGAGAAGCTT 1359
 |||||||
 QY 1306 TGTGCGACCTGCTGTGACGCTCAGTCACTTCTCTGAAAGCTTCTGAGCCCTTGACA 1365
 |||||||
 Db 1360 TGTGCGACCTGCTGTGACGCTCAGTCACTTCTCTGAAAGCTTCTGAGCCCTTGACA 1419
 |||||||
 QY 1366 TGTGCGACCTGCTGTGACGCTCAGTCACTTCTCTGAAAGCTTCTGAGCCCTTGACA 1425
 |||||||
 Db 1420 TGTGCGACCTGCTGTGACGCTCAGTCACTTCTCTGAAAGCTTCTGAGCCCTTGACA 1479
 |||||||
 QY 1426 TGTGCGACCTGCTGTGACGCTCAGTCACTTCTCTGAAAGCTTCTGAGCCCTTGACA 1485
 |||||||
 Db 1480 TGTGCGACCTGCTGTGACGCTCAGTCACTTCTCTGAAAGCTTCTGAGCCCTTGACA 1539
 |||||||
 QY 1486 TATGTGACGACGACATCCGACGATGAACAGGGTATGCTCTCTCTTCTGATCT 1545
 |||||||
 Db 1540 TATGTGACGACGACATCCGACGATGAACAGGGTATGCTCTCTCTTCTGATCT 1599
 |||||||
 QY 1546 TCTGTGCGACCTGCTGTGACGCTCAGTCACTTCTCTGAAAGCTTCTGAGCCCTTGACA 1605
 |||||||
 Db 1600 TCTGTGCGACCTGCTGTGACGCTCAGTCACTTCTCTGAAAGCTTCTGAGCCCTTGACA 1659
 |||||||
 QY 1606 ATCAGTGAATTTGATCCATAAACCTCCCTGTCACATCTTCCCAATGGGGAATGATCT 1665
 |||||||
 Db 1660 ATCAGTGAATTTGATCCATAAACCTCCCTGTCACATCTTCCCAATGGGGAATGATCT 1719
 |||||||
 QY 1666 TTCACCAAGAGCTCCACGACATTTTCCACAGAGATGC 1703
 |||||||
 Db 1720 TTCACCAAGAGCTCCACGACATTTTCCACAGAGATGC 1757
 |||||||

RESULT 11
 AF149783 1333 bp mRNA linear PRI 02-JUL-2001

LOCUS Homo sapiens L-selectin ligand sulfotransferase mRNA, complete cds.
 DEFINITION AF149783
 ACCESSION AF149783
 VERSION AF149783.1 GI:13897503
 KEYWORDS

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 1333)
 Yeh, J.C., Hiraoka, N., Petryniak, B., Nakayama, J., Elles, L.G.,
 Mammalia: Eutheria; Primates; Catarrhini; Hominoidea; Homo.

AUTHORS Yeh, J.C., Hiraoka, N., Petryniak, B., Nakayama, J., Elles, L.G.,
 Raduka, D., Hinds, J., O., Marth, J.D., Lowe, J.B., and Fukuda, M.
 Novel sulfated lymphocyte homing receptors and their control by a
 corel extension beta 1,3-N-acetylglucosaminyltransferase

JOURNAL Cell 105 (7), 957-969 (2001)
 MEDLINE 21332592
 PUBMED 11439191

REFERENCE 2 (bases 1 to 1333)
 Hiraoka, N. and Fukuda, M.
 Direct Submission
 Submitted (10-MAY-1999) Glycobiology, The Burnham Institute, 10901
 North Torrey Pines Road, La Jolla, CA 92037, USA

FEATURES
 SOURCE Location/Qualifiers
 1..1333
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 93..1253
 /note="LST"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase"
 /protein_id="AAK48417.1"
 /db_xref="GI:13897504"

CDS
 /translation="MLPKKKLLFLYSOMAILALFPHWYSHNISI.SMKAGPERKH
 VLVLSWRSGSEVFGULFGQHPDYFLMEPWHYMTFKOSTAMLMNAYDILLRAVF
 LQDMSEVDAYMEPGRSSILFQENSRALCSAPACDIIIPDELIIPAHCHLCSODP
 FEVEKACRSYSHVLEKEVFENLISYPLLKPSLMLHIVHLVDRBRAVRSRERIK

BASE COUNT 286 a 393 c 336 g 318 t
 ORIGIN
 GDLMDISRIYWGHEOKLKEDQPYVWQVICOQSOLEIYKTIQSLPALORRYLLVRY
 EDLAPYVAGTSMRYEFYLEFLPHLDITWVINITRGMGMGHAFHTNARALVNSQAW
 RMSLPYKVSLSOKACGDANMLLRYRVRSRQBRNLLDLSTWYPEQIH"

Query Match 61.8%; Score 1262; DB 9; Length 1333;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 107 AAGCTTCCACTTCAGACATGCTACTGCTTAAAAAATGAAGCTCTCTGTTTGTG 166
 |||||||
 Db 72 AAGCTTCCACTTCAGACATGCTACTGCTTAAAAAATGAAGCTCTCTGTTTGTG 131
 |||||||
 QY 167 GTTCCGATGGGACATCTTGCTATCTTCCACATGTACAGACACATCAGCTCC 226
 |||||||
 Db 132 GTTCCGATGGGACATCTTGCTATCTTCCACATGTACAGACACATCAGCTCC 191
 |||||||
 QY 227 CTGTCTATGAAGGACACGCCGAGCGCATGACGCTGCTCTCTCTGCGCTCT 286
 |||||||
 Db 192 CTGTCTATGAAGGACACGCCGAGCGCATGACGCTGCTCTCTCTGCGCTCT 251
 |||||||
 QY 287 GGCCTCTTTTGTGGGACACTTTTGGGACAGACCCAGATGTTTCTACTGATGAG 346
 |||||||
 Db 252 GGCCTCTTTTGTGGGACACTTTTGGGACAGACCCAGATGTTTCTACTGATGAG 311
 |||||||
 QY 347 CCGGCTGGGACGCTGGATGACCTTCAAGACAGACCGGCTGGATGCTCAGATGCT 406
 |||||||
 Db 312 CCGGCTGGGACGCTGGATGACCTTCAAGACAGACCGGCTGGATGCTCAGATGCT 371
 |||||||
 QY 407 GTGGGGATCTGATACGGGCGCTCTTCTGTGCGACATGACGCTTCTGATGCTTAC 466
 |||||||
 Db 372 GTGGGGATCTGATACGGGCGCTCTTCTGTGCGACATGACGCTTCTGATGCTTAC 431
 |||||||
 QY 467 GAACCTGTCCTCCGAGACATGACGCTCTTTCAGTGGGAGAACGCCGCGCTGTCT 526
 |||||||
 Db 432 GAACCTGTCCTCCGAGACATGACGCTCTTTCAGTGGGAGAACGCCGCGCTGTCT 491
 |||||||
 QY 527 TCTGACCTGCTGTCGATCATGCCACAAATGAAATCAATCCCGGCGCTCAGTGGAG 586
 |||||||
 Db 492 TCTGACCTGCTGTCGATCATGCCACAAATGAAATCAATCCCGGCGCTCAGTGGAG 551
 |||||||
 QY 587 CTCCTGTGACATCAACAGCCCTTGTGAGTGTGAGAGAGGCTGCGCTCTACAGCCAC 646
 |||||||
 Db 552 CTCCTGTGACATCAACAGCCCTTGTGAGTGTGAGAGAGGCTGCGCTCTACAGCCAC 611
 |||||||
 QY 647 GTGTGCTCAAGAGAGTCCGCTTCTTCAACTGCAAGTCCCTCTACCCGCTGTAAGAC 706
 |||||||
 Db 612 GTGTGCTCAAGAGAGTCCGCTTCTTCAACTGCAAGTCCCTCTACCCGCTGTAAGAC 671
 |||||||
 QY 707 CCGTCCGACACGTCGATATGTGACACTGTGCGGGAGACCCCGGCGCGTTCGCTCC 766
 |||||||
 Db 672 CCGTCCGACACGTCGATATGTGACACTGTGCGGGAGACCCCGGCGCGTTCGCTCC 731
 |||||||
 QY 767 CGAAGACGACAAAGGAGATCTCATGTTTACAGTCCATGTTGATGGGACAGATGAG 826
 |||||||
 Db 732 CGAAGACGACAAAGGAGATCTCATGTTTACAGTCCATGTTGATGGGACAGATGAG 791
 |||||||
 QY 827 CAAAAACTCAAGAGAGAGACCAACCTTACTATGTATGATGACAGTCACTGTCCAAAGCCAG 886
 |||||||
 Db 792 CAAAAACTCAAGAGAGAGACCAACCTTACTATGTATGATGACAGTCACTGTCCAAAGCCAG 851
 |||||||
 QY 887 CTGAGATCTCAAGAGACATCCAGTCTTGTGCGCAAGGCGCTGCGAGAACCGTACTGCTT 946
 |||||||
 Db 852 CTGAGATCTCAAGAGACATCCAGTCTTGTGCGCAAGGCGCTGCGAGAACCGTACTGCTT 911
 |||||||
 QY 947 GTGGCTATGAGAGACTGCTGAGAGCCCTGTGCGCCAGACTTCCGAAATGATGAATTC 1006
 |||||||
 Db 912 GTGGCTATGAGAGACTGCTGAGAGCCCTGTGCGCCAGACTTCCGAAATGATGAATTC 971
 |||||||
 QY 1007 GTGGATGTAATTCCTGCGCCATCTTCAAGCTGCGGTGATCAATCACCAGGAGCAAG 1066
 |||||||
 Db 972 GTGGATGTAATTCCTGCGCCATCTTCAAGCTGCGGTGATCAATCACCAGGAGCAAG 1031
 |||||||

QY 1067 GGCATGGGTGACACGCTTTCACACAATGCGAGGATGCGCTTAATGTCTCCAGGCT 1126
 DB 1032 GGCATGGGTGACACGCTTTCACACAATGCGAGGATGCGCTTAATGTCTCCAGGCT 1091
 QY 1127 TGGCGGTGCTTGGCCCTATGAAAAGGTTTCGACTTCAGAAAAGCTTGGCGATGCC 1186
 DB 1092 TGGCGGTGCTTGGCCCTATGAAAAGGTTTCGACTTCAGAAAAGCTTGGCGATGCC 1151
 QY 1187 ATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAACAAGACAGAAAAGCTTGTGCTG 1246
 DB 1152 ATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAACAAGACAGAAAAGCTTGTGCTG 1211
 QY 1247 GATCTTGTCTACCTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGGTTGAGAAGCTTT 1306
 DB 1212 GATCTTGTCTACCTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGGTTGAGAAGCTTT 1271
 QY 1307 GCGGCAACGCGGTGAGCGCTGACGCTTCTCTGAATGCTCTGAGCTTGGCTACAT 1366
 DB 1272 GCTGCCACCTGGGTGTCAGCTCAGTCACTTCTCTGAATGCTCTGAGCTTGGCTACAT 1331
 QY 1367 CT 1368
 DB 1332 CT 1333
 RESULT 12
 AR310485 AR310485 2065 bp DNA linear PAT 12-JUN-2003
 LOCUS AR310485 Sequence 26 from patent US 6558935.
 DEFINITION AR310485
 ACCESSION AR310485
 VERSION AR310485.1 GI:31703448
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2065)
 AUTHORS Tang,Y.T., Corley,N.C., Guegler,K.J., Baughn,M.R., Lal,P., Yue,H.,
 TITLE Human Transferrin Proteins
 JOURNAL Patent: US 6558935-A 26 06-MAY-2003;
 FEATURES
 source location/Qualifiers
 1..2065
 BASE COUNT 466 a 575 c 491 g 533 t
 ORIGIN
 Query Match 56.4%; Score 1153; DB 6; Length 2065;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

DB 452 TGTGGGATCTGATACGGGCGCTCTCTTGTGCGACATGAGCGCTTGTGATCTACAT 511
 QY 466 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAGAACAGCGGCGCTGTG 525
 DB 512 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAGAACAGCGGCGCTGTG 571
 QY 526 TTCTGCACCTGCTGTGACATCATCCACAAGATGAATATATCCCGGGCTCACTGAC 585
 DB 572 TTCTGCACCTGCTGTGACATCATCCACAAGATGAATATATCCCGGGCTCACTGAC 630
 QY 586 GCTTCGTGTGAGTCAACAGCCCTTTG-AGTGGTGGAGAAAGCCCTGCTCTACAGCC 644
 DB 631 GCTTCGTGTGAGTCAACAGCCCTTTGAGTGGAGAAAGCCCTGCTCTACAGCC 690
 QY 645 AGTGGTGTCAAGAGAGTCCGCTTCTCAACTGAGAGTCCCTACCCGCTCTGAAG 704
 DB 691 AGTGGTGTCAAGAGAGTCCGCTTCTCAACTGAGAGTCCCTACCCGCTCTGAAG 750
 QY 705 ACCCTGCTCAACCTGATATGTCGACCTGGTCCGGGACCCCGGGCGTGTCCGTT 764
 DB 751 ACCCTGCTCAACCTGATATGTCGACCTGGTCCGGGACCCCGGGCGTGTCCGTT 810
 QY 765 CCGGAGAACGCAAAAGGAGATCATGATTGACAGTGGCATTTGTATGGGAGACATG 824
 DB 811 CCGGAGAACGCAAAAGGAGATCATGATTGACAGTGGCATTTGTATGGGAGACATG 870
 QY 825 AGCAAAACCTCAAGAGAGAGACCAACCCCTACTATGTATGATGATGATGCAAGCC 884
 DB 871 AGCAAAACCTCAAGAGAGAGACCAACCCCTACTATGTATGATGATGATGCAAGCC 930
 QY 885 AGCTGAGATCTTCAAGAGACCATTCAGTCTTGGCCCAAGGCGCTGCAAGAAAGCTTAC 944
 DB 931 AGCTGAGATCTTCAAGAGACCATTCAGTCTTGGCCCAAGGCGCTGCAAGAAAGCTTAC 990
 QY 945 TTGTGCGATGAGAGAGTCTGAGAGCCCTGAGAGCCCTGAGAGTCCGCAATGATGAT 1004
 DB 991 TTGTGCGATGAGAGAGTCTGAGAGCCCTGAGAGCCCTGAGAGTCCGCAATGATGAT 1050
 QY 1005 TCGTGGGATTTGATTTCTGCCCCATCTTCAGACCTGGGTGATTAACATCACCGAGCA 1064
 DB 1051 TCGTGGGATTTGATTTCTGCCCCATCTTCAGACCTGGGTGATTAACATCACCGAGCA 1110
 QY 1065 AGGGCATGGGTGACACAGCTTTTCCACACAATGCCAGGATGCCCTTAATGTCTCCAG 1124
 DB 1111 AGGGCATGGGTGACACAGCTTTTCCACACAATGCCAGGATGCCCTTAATGTCTCCAG 1170
 QY 1125 CTGGCGCTGCTTGTGCTTATGAAAAGTTTCTGAGCTTCAAGAAAGCTGCGGAGAT 1184
 DB 1171 CTGGCGCTGCTTGTGCTTATGAAAAGTTTCTGAGCTTCAAGAAAGCTGCGGAGAT 1230
 QY 1185 CCATGAATTTGCTGGGCTACCGCCACGTCAGATCTGACAAAGAACAGAAACCTGTGC 1244
 DB 1231 CCATGAATTTGCTGGGCTACCGCCACGTCAGATCTGACAAAGAACAGAAACCTGTGC 1290
 QY 1245 TGGATCTTCTGCTTACCTGAGCTGCTCCGAGCAATCCACTAAGAGGTTGGAAGCT 1304
 DB 1291 TGGATCTTCTGCTTACCTGAGCTGCTCCGAGCAATCCACTAAGAGGTTGGAAGGCT 1350
 QY 1305 TTGCTGCCACCTGGTGTACAGCTCACTCACTTCTCTGATGCTTGTGAGCCCTGCTTAC 1364
 DB 1351 TTGCTGCCACCTGGTGTACAGCTCACTCACTTCTCTGATGCTTGTGAGCCCTGCTTAC 1410
 QY 1365 ATCTGTAGACCTTAACATATGCTGTGGGTATACACTGAGTGTGATGTGTCCACAC 1424
 DB 1411 ATCTGTAGACCTTAACATATGCTGTGGGTATACACTGAGTGTGATGTGTGTCCACAC 1470
 QY 1425 GTGCTTAACAGAGAGACTTTGTGTGCATGCTGTGTGTATAGAAAACAACACTGGGAAAC 1484
 DB 1471 GTGCTTAACAGAGAGACTTTGTGTGCATGCTGTGTGTATAGAAAACAACACTGGGAAAC 1530
 QY 1485 TTATGTAGAGAGACATCCACAGTGAAGAAAGAGGATTTGCTTCTTTCTGTGATC 1544
 DB 1531 TTATGTAGAGAGACATCCACAGTGAAGAAAGGATTTGCTTCTTTCTGTGATC 1590

```

QY 1545 TTCTGCTGGGAGAGCTTACAGAGCTTTGGGCTGGAGGCTATTAAAGCAGACAG 1604
DB 1591 TTCTGCTGGGAGAGCTTACAGAGCTTTGGGCTGGAGGCTATTAAAGCAGACAG 1650
QY 1605 TATAGTGGAGTATATCCATTAACCTCCCTGTCACATCTTGCCCAATGGGAAATGATC 1664
DB 1651 TATAGTGGAGTATATCCATTAACCTCCCTGTCACATCTTGCCCAATGGGAAATGATC 1710
QY 1665 TTTCACCAAGAGCTCCACACATTTTCCACAGAGATGC 1703
DB 1711 TTTCACCAAGAGCTCCACACATTTTCCACAGAGATGC 1749

RESULT 13
BD124832
LOCUS BD124832 877 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD124832.1 GI:23219777
VERSION JP 2002017375-A/263
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002017375-A 263 22-JAN-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/263
PD 22-JAN-2002
PE 07-JUL-2000 JP 2000253172
PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
PI ISHII,
PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI,HISASHI KOGA
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
10,
PC C12P21/02,C1201/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT source 1..877
/organism="Homo sapiens (human)".
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 183 a 266 c 221 g 203 t 4 others
ORIGIN
Query Match 34.0%; Score 695; DB 6; Length 877;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 CAAGGCTCTCCACTTACAGACATGCTACTGCTTAAATAATGAAGCTCCTGCTTCT 165
DB 47 CAAGGCTCTCCACTTACAGACATGCTACTGCTTAAATAATGAAGCTCCTGCTTCT 106
QY 166 GGTTCACAGATGGCCATCTGGCTATTCTTCCACATGTACAGCCACAACATCAGCTC 225
DB 107 GGTTCACAGATGGCCATCTGGCTATTCTTCCACATGTACAGCCACAACATCAGCTC 166
QY 226 CCGTCTATGAAGCAGACGCCGAGCGATGACAGTGTGCTGCTTCTGCTGCGCTC 285
DB 167 CCGTCTATGAAGCAGACGCCGAGCGATGACAGTGTGCTGCTTCTGCTGCGCTC 226

FEATURES
source
Location/Qualifiers
1..877
/organism="Homo sapiens (human)".
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 183 a 266 c 221 g 203 t 4 others
ORIGIN
Query Match 34.0%; Score 695; DB 6; Length 877;
Best Local Similarity 100.0%; Pred. NO. 0;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 286 TGCGTCTTCTTTTGTGGGGAGCTTTTGGGACACACAGATGTTTCTACATGATGA 345
DB 227 TGCGTCTTCTTTTGTGGGGAGCTTTTGGGACACACAGATGTTTCTACATGATGA 286
QY 346 GCCCGCTGGCAGCTGTGATGACCTTCAACAGACAGCCGCTGATGTGCAATGGC 405
DB 287 GCCCGCTGGCAGCTGTGATGACCTTCAACAGACAGCCGCTGATGTGCAATGGC 346
QY 406 TGTGGGGATGTGATGAGGGGCGTCTTGTGGAGACATGAGCGCTTGTGATGCTACAT 465
DB 347 TGTGGGGATGTGATGAGGGGCGTCTTGTGGAGACATGAGCGCTTGTGATGCTACAT 406
QY 466 GGAACCTGTGTCGCCGAGACAGTCCAGGCTTTTCAGTGGAGAAACACCGGCGCTGTG 525
DB 407 GGAACCTGTGTCGCCGAGACAGTCCAGGCTTTTCAGTGGAGAAACACCGGCGCTGTG 466
QY 526 TTTCGACCTGCTGCTGTGACATCACCACAAGATGAATCATCCCGGCGCTCACTGAG 585
DB 467 TTTCGACCTGCTGCTGTGACATCACCACAAGATGAATCATCCCGGCGCTCACTGAG 526
QY 586 GCTCCTGTGAGTCAACAGCCCTTTGAGGTGTGAGAGAGCGCTGCTCTACAGCA 645
DB 527 GCTCCTGTGAGTCAACAGCCCTTTGAGGTGTGAGAGAGCGCTGCTCTACAGCA 586
QY 587 CGTGCTGTCAAGAGAGTGCCTTCTTCAACCTGACGCTTACCCGCTGTGAAAGA 705
DB 646 CGTGCTGTCAAGAGAGTGCCTTCTTCAACCTGACGCTTACCCGCTGTGAAAGA 587
QY 647 CCCCTCCTCAACCTGATATGACGCTGCTGACGCTGCTGCTGCTGCTGCTGCTGCTG 706
DB 707 CCGAGAACGCAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 741

RESULT 14
BD126950
LOCUS BD126950 877 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD126950.1 GI:23221895
VERSION JP 2002017375-A/2381.
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
Primer for synthesizing full-length cDNA and use thereof
Patent: JP 2002017375-A 2381 22-JAN-2002;
HELIX RESEARCH INSTITUTE
OS Homo sapiens (human)
PN JP 2002017375-A/2381
PD 22-JAN-2002
PE 07-JUL-2000 JP 2000253172
PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
PI ISHII,
PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
SHINICHI KOJIMA,
PI TETSUJI OTSUKI,HISASHI KOGA
PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
10,
PC C12P21/02,C1201/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
Primer for synthesizing full-length cDNA and use thereof FH Key
Location/Qualifiers
FT source 1..877
/organism="Homo sapiens (human)".
Location/Qualifiers

```


source 1. 877
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 183 a 266 c 221 g 203 t 4 others
 ORIGIN

Query Match 34.0%; Score 695; DB 6; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 CAAGCTTCCACTTACGACATGCTACGCTAAATAAATCACTCTCTGTTCT 165
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 47 CAAAGCTTCCACTTACGACATGCTACGCTAAATAAATCACTCTCTGTTCT 106
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 166 GGTGTTCCAGATGGCCATCTGTGCTATCTTCCACATGTACAGCCACAACTCAGCTC 225
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 107 GGTGTTCCAGATGGCCATCTGTGCTATCTTCCACATGTACAGCCACAACTCAGCTC 166
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 226 CCGTGTATGAAGGACACAGCCGAGCGCATGACGCTGCTGTTCTCTCGGCGCTC 285
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 167 CCGTGTATGAAGGACACAGCCGAGCGCATGACGCTGCTGTTCTCTCGGCGCTC 226
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 286 TGAGCTCTCTTGTGAGGAGCTTTTGGGACAGCCAGATGTTTCTACCTGATGA 345
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 227 TGAGCTCTCTTGTGAGGAGCTTTTGGGACAGCCAGATGTTTCTACCTGATGA 286
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 346 GCCCGCCTGGACAGCTGTGATGACCTTCAAGCAGAGACCCGCTGATGCTGACATGGC 405
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 287 GCCCGCCTGGACAGCTGTGATGACCTTCAAGCAGAGACCCGCTGATGCTGACATGGC 346
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 406 TGTGGGGGATGATACGAGGCGCTCTTGTGCGACATGAGGCTTTGATGCTACAT 465
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 347 TGTGGGGGATGATACGAGGCGCTCTTGTGCGACATGAGGCTTTGATGCTACAT 406
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 466 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGAGAGACCGCGGCGCTG 525
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 407 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGAGAGACCGCGGCGCTG 466
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 526 TTCTGCACCTGCTGTGATCATCTCCACAAATGAAATCATCCCCGGGCTCAGTCAG 585
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 467 TTCTGCACCTGCTGTGATCATCTCCACAAATGAAATCATCCCCGGGCTCAGTCAG 526
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 586 GCTCCGTGCAAGTCAACACACCCCTTTGAGGTGGTGGAGAGGCTGCCGCTCCTACAGCA 645
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 527 GCTCCGTGCAAGTCAACACACCCCTTTGAGGTGGTGGAGAGGCTGCCGCTCCTACAGCA 586
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 646 CGTGTGCTCAAGAGAGTGGCTTTCTTCAACCTGCAGTCCCTTACCCGCTGCTGAAGA 705
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 587 CGTGTGCTCAAGAGAGTGGCTTTCTTCAACCTGCAGTCCCTTACCCGCTGCTGAAGA 646
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 706 CCCCTCCCTCAACCTGCATATGCTGCAGCTGTCCGGACCCCGGGCGTGTCCGTTTC 765
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 647 CCCCTCCCTCAACCTGCATATGCTGCAGCTGTCCGGACCCCGGGCGTGTCCGTTTC 706
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 766 CCGAGAACCCCAAAAGGAGATCTCATGATGACA 800
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 707 CCGAGAACCCCAAAAGGAGATCTCATGATGACA 741
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 15
 AX381256
 LOCUS AX381256 517 bp DNA 11near PAT 18-MAR-2002
 DEFINITION Sequence 194 from Patent WO0212280.
 ACCESSION AX381256
 VERSION AX381256.1 GI:19576075
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 Pyle, R.A., Xu, J. and Secrist, H.

TITLE Compositions and methods for the therapy and diagnosis of colon
 JOURNAL cancer
 Patent: WO 0212280-A 194 14-FEB-2002;
 CORIXA CORPORATION (US)
 FEATURES location/Qualifiers
 source 1. 517
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

BASE COUNT 121 a 130 c 124 g 142 t
 ORIGIN

Query Match 19.7%; Score 403; DB 6; Length 517;
 Best Local Similarity 100.0%; Pred. No. 3.5e-228;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCTTTCGCGCCACCTGGTGTGACGCTCAGCTCTTCTGTGAATGCTGTAGCCCTGC 1360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 13 GGCTTTCGCGCCACCTGGTGTGACGCTCAGCTCTTCTGTGAATGCTGTAGCCCTGC 72
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1361 CTACATCTGTGACCTTAACTACATGCTGTGGGTATCACACTGAGTGTGATGTC 1420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 73 CTACATCTGTGACCTTAACTACATGCTGTGGGTATCACACTGAGTGTGATGTC 132
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1421 ACAAGTGTCAACAGAGAGACTTGTGTCCATGCTGTGTGTATAGAAAAACACTGGGG 1480
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 133 ACAAGTGTCAACAGAGAGACTTGTGTCCATGCTGTGTGTATAGAAAAACACTGGGG 192
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1481 AACCTTATGTGACAGACACATCCACAGAGTGAAGAGATGCTCTTCTTCTTCTT 1540
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 193 AACCTTATGTGACAGACACATCCACAGAGTGAAGAGATGCTCTTCTTCTTCTT 252
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1541 GATCTTCTGTGTGGGACAGCTTCAAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1600
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 253 GATCTTCTGTGTGGGACAGCTTCAAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 312
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1601 ACAGTATCACTGGAATTTGATCCATAACCTCCCTGTCCACATCTTCCCAATGGGAATG 1660
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 313 ACAGTATCACTGGAATTTGATCCATAACCTCCCTGTCCACATCTTCCCAATGGGAATG 372
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1661 GATCTTTCACCAAGAGCTCAGCAGATTTTCCAGAGATGC 1703
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 373 GATCTTTCACCAAGAGCTCAGCAGATTTTCCAGAGATGC 415
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Search completed: August 15, 2003, 06:52:27
 Job time : 7290 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2003, 01:05:28 ; Search time 542 Seconds
(without alignments)
10175.188 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattccattgtgttggtta.....tgggatacttaaaaaaaaaa 2043

Scoring table:

OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1856220

Minimum DB seq length: 300

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database: N_Geneseq_19Jun03.*

1: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT.*
2: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT.*
3: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT.*
4: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT.*
5: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT.*
6: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT.*
7: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT.*
8: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT.*
9: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT.*
10: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT.*
11: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT.*
12: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT.*
13: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT.*
14: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT.*
15: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT.*
16: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT.*
17: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT.*
18: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT.*
19: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT.*
20: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT.*
21: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT.*
22: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT.*
23: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT.*
24: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT.*
25: /SIDSI/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	ID	Description
1	1547	75.7	1979	22 AAK94229
2	1547	75.7	2032	20 AAZ20792
3	1262	61.8	1333	24 AAS16947
4	1153	56.4	2065	21 AAZ94211
5	695	34.0	877	22 AAK91803
6	695	34.0	877	22 AAK93921
7	403	19.7	505	24 ABV89280
8	403	19.7	517	24 AAK54724

9	306	15.0	2988	21 AAK76156	Human OREF ORF1711
10	114	5.6	389	24 AAK54794	Human colon cancer
11	105	5.1	548	22 AAK92588	Human cDNA 3'-end
12	59	2.9	1647	24 AAD24670	Human drug metabol
13	59	2.9	1694	22 AAD02700	Human glycosyl sul
14	59	2.9	2044	22 AAD02699	Human glycosyl sul
15	59	2.9	2170	22 AAD02698	Human glycosyl sul
16	59	2.9	2544	24 AAB99505	Human corneal N-ac
17	59	2.9	48436	24 ABB88533	Human corneal N-ac
18	59	2.9	160552	22 AAD02697	Human glycosyl sul
19	27	1.3	1926	20 AAZ20793	Mouse glycosyl sul
20	23	1.1	1937	24 AAS16948	Mouse glycosyl sul
21	23	1.1	1989	22 AAD02696	Mouse glycosyl sul
22	22	1.0	10004	22 ABA14483	Mouse glycosyl sul
23	21	1.0	10246	22 ABA14485	Human nervous syst
24	21	1.0	354	22 AAF65813	Human nervous syst
25	20	1.0	370	22 AAF65341	Novel human polynu
26	20	1.0	381	22 AAF67389	Novel human polynu
27	20	1.0	398	25 ABA45934	Bovine EST associa
28	20	1.0	3763	19 AAV58194	Human myosin I-cha
29	20	1.0	3763	23 AAS69872	DNA encoding novel
30	20	1.0	5814	24 ABB87855	Human ovary specif
31	20	1.0	6147	25 ACC46453	Human dlthp protei
32	20	1.0	9516	22 AAK73555	Human immune/haema
33	20	1.0	9516	22 AAK81394	Human immune/haema
34	20	1.0	13192	22 AAK73556	Human immune/haema
35	20	1.0	13192	22 AAK81396	Human immune/haema
36	20	1.0	33959	22 AAK78275	Human immune/haema
37	20	1.0	180557	24 ABB85750	Human BAC clone RP
38	20	1.0	349980	22 AAF86431	Pyrococcus abyssi
39	19	0.9	341	24 ABL85863	Human ovarian canc
40	19	0.9	341	24 ABL85910	Human ovarian canc
41	19	0.9	348	22 AAF67078	Novel human polynu
42	19	0.9	349	22 AAS63985	Human prostate CDN
43	19	0.9	349	24 ABL95356	Human prostate-spe
44	19	0.9	349	25 ACA59793	Prostate cancer th
45	19	0.9	353	22 AAF66848	Novel human polynu

ALIGNMENTS

RESULT 1

AAK94229 standard; cDNA: 1979 BP.

AAK94229;

06-NOV-2001 (first entry)

Human full-length cDNA, SEQ ID NO: 2816.

Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

Homo sapiens.

EP1130094-A2.

05-SEP-2001.

07-JUL-2000; 2000EP-0114089.

08-JUL-1999; 99JP-0194486.

11-JAN-2000; 2000JP-0118774.

02-MAY-2000; 2000JP-0185765.

(HELI-) HELIX RES INST.

Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

WPI; 2001-524255/58.

P-PSDB; AAM93309.

XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
XX
PS Claim 8; SEQ ID NO 2816; 1380bp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SO Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other;

Query Match 75.7%; Score 1547; DB 22; Length 1979;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAAGGCTTCACATTAGACACAAATGCTACTGCTTAAATAAAGAGCTGCTGTTCT 165
DB 47 CAAGGCTTCACATTAGACACAAATGCTACTGCTTAAATAAAGAGCTGCTGTTCT 106
QY 166 GATTTCACAGATGGCATCTGCTATCTTCCACATGATACAGCACAACATCAGCTC 225
DB 107 GATTTCACAGATGGCATCTGCTATCTTCCACATGATACAGCACAACATCAGCTC 166
QY 226 CCGTGTATGAAGGACACAGCCGAGCGCATGACGCTGCTGCTGCTGCTGCTGCTGCT 285
DB 167 CCGTGTATGAAGGACACAGCCGAGCGCATGACGCTGCTGCTGCTGCTGCTGCTGCT 226
QY 286 TGCTCTCTCTTCTTGTGGGACGCTTGTGGGACGACCCAGATGTTTCTACTGATGGA 345
DB 227 TGCTCTCTCTTCTTGTGGGACGCTTGTGGGACGACCCAGATGTTTCTACTGATGGA 286
QY 346 GCCCGCTGACAGCTGTGATGACCTTCAAGACAGAGACAGCCGCTGATGCTGACATGGC 405
DB 287 GCCCGCTGACAGCTGTGATGACCTTCAAGACAGAGACAGCCGCTGATGCTGACATGGC 346
QY 406 TGTGCGGATCTGATACGGGCGCTGCTTGTGCGACATGAGCGCTTGTGATGCTGAT 465
DB 347 TGTGCGGATCTGATACGGGCGCTGCTTGTGCGACATGAGCGCTTGTGATGCTGAT 406
QY 466 GGAACCTGATCCCGGAGACAGTCCAGCCCTTTCAGTGGGANAAGCCGGGCGCTGTG 525
DB 407 GGAACCTGATCCCGGAGAGAGTCCAGCCCTTTCAGTGGGANAAGCCGGGCGCTGTG 466
QY 526 TTCTGACCTGCTGATGATCATCCACAGATGAATCATCCCGGCGCTACTGACAG 585
DB 467 TTCTGACCTGCTGATGATCATCCACAGATGAATCATCCCGGCGCTACTGACAG 526
QY 586 GCTCCCTGATGATGACAGCCCTTGAAGTGTGGAAGAGCGCTGCTGCTGCTGCTGCT 645
DB 527 GCTCCCTGATGATGACAGCCCTTGAAGTGTGGAAGAGCGCTGCTGCTGCTGCTGCT 586
QY 646 CGTGTGCTCAAGAGAGTGGCTTTCACACCTGAGTCCCTTACCCGCTGCTGAAGA 705
DB 587 CGTGTGCTCAAGAGAGTGGCTTTCACACCTGAGTCCCTTACCCGCTGCTGAAGA 646
QY 706 CCCCTCCCTCAACCTGATGATGCTGACCTGCTGCGGAGCCCGGCGCTGCTGCTGCT 765
DB 647 CCCCTCCCTCAACCTGATGATGCTGACCTGCTGCGGAGCCCGGCGCTGCTGCTGCT 706
QY 766 CCGAGAGACGACAAAGGAGATCTCATGATGACAGCTGCGATGATGATGATGATGATGAT 825
DB 707 CCGAGAGACGACAAAGGAGATCTCATGATGACAGCTGCGATGATGATGATGATGATGAT 766

QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTCACTATGATGACAGCTATCTGCCAAAGCCA 885
DB 767 GCAAAAACCTCAAGAGAGAGACCAACCTCACTATGATGACAGCTATCTGCCAAAGCCA 826
QY 886 GCTGAGATCTAAGACCACTCCAGTCTTGGCCCAAGGCGCTCAGAAACGCTACCTGCT 945
DB 827 GCTGAGATCTAAGACCACTCCAGTCTTGGCCCAAGGCGCTCAGAAACGCTACCTGCT 886
QY 946 TGTGCGCTATGAGAGACTGCTGAGACCCCTGCTGCGCCAGACTTCCGATGATGAT 1005
DB 887 TGTGCGCTATGAGAGACTGCTGAGACCCCTGCTGCGCCAGACTTCCGATGATGAT 946
QY 1006 CGTGGATGGAATTTCTTGGCCCATTTTCAGACCTGAGCTGATATACATCAGCGAGCA 1065
DB 947 CGTGGATGGAATTTCTTGGCCCATTTTCAGACCTGAGCTGATATACATCAGCGAGCA 1006
QY 1066 GGGCATGGGAGACAGCTTTCACACAAATGACAGGAGTGGCTTAATGTCGCCAGGC 1125
DB 1007 GGGCATGGGAGACAGCTTTCACACAAATGACAGGAGTGGCTTAATGTCGCCAGGC 1066
QY 1126 TTGCGCTGCTGCTTTCCTCATGAAAAGTTCCTGACCTTCAGAAAAGCTGCTGCGATGC 1185
DB 1067 TTGCGCTGCTGCTTTCCTCATGAAAAGTTCCTGACCTTCAGAAAAGCTGCTGCGATGC 1126
QY 1186 CATGAATTTGCTGCGCTACCGCCAGCTCAGATCTGAACAAGAACAGAAACCTGTTGCT 1245
DB 1127 CATGAATTTGCTGCGCTACCGCCAGCTCAGATCTGAACAAGAACAGAAACCTGTTGCT 1186
QY 1246 GGAATTTGCTGCTACCTGAGCTGCTGCTGAGAAATCCACTAGAGGATGAGAAAGCTT 1305
DB 1187 GGAATTTGCTGCTACCTGAGCTGCTGCTGAGAAATCCACTAGAGGATGAGAAAGCTT 1246
QY 1306 TGCTGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1365
DB 1247 TGCTGACACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1306
QY 1366 TCTCTGACCTTAACTACATGCTGTGGGATACACTGATGATGATGATGATGATGATGAT 1425
DB 1307 TCTCTGACCTTAACTACATGCTGTGGGATACACTGATGATGATGATGATGATGATGAT 1366
QY 1426 TGCTGACGAGAAAGGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1485
DB 1367 TGCTGACGAGAAAGGAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1426
QY 1486 TATGTAGAGACGACATCCACAGAGTAAAGAGGATGCTGCTTCTTCTTCTGATCT 1545
DB 1427 TATGTAGAGACGACATCCACAGAGTAAAGAGGATGCTGCTTCTTCTTCTGATCT 1486
QY 1546 TCTGTGCTGCGGAGACTTCAAGAGACTTGTGGGCTGAGAGCCATTAACAGACAGCACT 1605
DB 1487 TCTGTGCTGCGGAGACTTCAAGAGACTTGTGGGCTGAGAGCCATTAACAGACAGCACT 1546
QY 1606 ATCAGTGAATGATCATTAACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1665
DB 1547 ATCAGTGAATGATCATTAACCTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1606
QY 1666 TTACCAAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
DB 1607 TTACCAAAAGAGCTCACAGCATTTTCCACAGAGATGC 1644

RESULT 2
AAZ20792
ID AAZ20792 standard; DNA; 2032 BP.
XX
XX AAZ20792;
XX
XX
XX 08-DEC-1999 (first entry)
XX
XX Human glycosyl sulfotransferase-3 coding sequence.
DE Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW selectin binding interaction; inflammation; lymphocyte homing; human;

KW secondary lymph organ; ss.
 XX Homo sapiens.
 XX MO9949018-A1.
 XX 30-SEP-1999.
 XX 26-FEB-1999; 99MO-US04316.
 XX 20-MAR-1998; 98US-0045284.
 PR 12-NOV-1998; 98US-0190911.
 XX (REGC) UNIV CALIFORNIA.
 PA (SYNT) SYNTEX USA INC.
 XX Blstrup A, Rosen SD, Tangemann K, Hemmerich S;
 DR WPI: 1999-580442/49.
 DR P-PSDB; AAY39918.
 XX Human and murine glycosyl sulfotransferase 3 and related
 PT polynucleotides
 XX
 PS Claim 4; Fig 1; 59pp; English.
 CC This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.
 CC
 XX Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other;
 SQ
 Query Match 75.7%; Score 1547; DB 20; Length 2032;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 586 GCTCTGTGACATCAGACCCCTTTGAGGTGGTGGAGAGGCTGCCGCTCTACAGCA 645
 |||||||
 Db 640 GCTCTGTGACATCAGACCCCTTTGAGGTGGTGGAGAGGCTGCCGCTCTACAGCA 699
 QY 646 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAGA 705
 |||||||
 Db 700 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAGA 759
 QY 706 CCCCTCCCTCAACCTGATATGCGACACCTGCTCCGGAGACCCCGGGCCGCTCCGCTC 765
 |||||||
 Db 760 CCCCTCCCTCAACCTGATATGCGACACCTGCTCCGGAGACCCCGGGCCGCTCCGCTC 819
 QY 766 CCGAGAACGCACAAAGGAGATCTCATGATGACAGTGCATTTGTGATGGGCGAGATGA 825
 |||||||
 Db 820 CCGAGAACGCACAAAGGAGATCTCATGATGACAGTGCATTTGTGATGGGCGAGATGA 879
 QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTACTATGTGATGAGGTCACTGCCAAAGCA 885
 |||||||
 Db 880 GCAGAAACTCAAGAGAGAGACCAACCTACTATGTGATGAGGTCACTGCCAAAGCA 939
 QY 886 GCTGAGATCTACAGACATCCAGTCCCTTCCCAAGGCCCTGAGAGAGCTACCTGCT 945
 |||||||
 Db 940 GCTGAGATCTACAGACATCCAGTCCCTTCCCAAGGCCCTGAGAGAGCTACCTGCT 999
 QY 946 TGTGCGTATGAGAGACCTGAGCTCGAGCCCTGTGGCCAGACTTCCGAAATGTATGAAT 1005
 |||||||
 Db 1000 TGTGCGTATGAGAGACCTGAGCTCGAGCCCTGTGGCCAGACTTCCGAAATGTATGAAT 1059
 QY 1006 CGTGGATTTGAATTTCTTGGCCCATCTTACAGACTGGGTGATTAACATACCCAGGCA 1065
 |||||||
 Db 1060 CGTGGATTTGAATTTCTTGGCCCATCTTACAGACTGGGTGATTAACATACCCAGGCA 1119
 QY 1066 GGGCATGGGTGACACAGCTTCCACACAAATGGCAGGATGCCCTTATGTCTCCAGGC 1125
 |||||||
 Db 1120 GGGCATGGGTGACACAGCTTCCACACAAATGGCAGGATGCCCTTATGTCTCCAGGC 1179
 QY 1126 TTGGCGGTGCTTTTGGCCCATATGAAAGTTTCTGACTTCAGAAAGCCTGTGGCATGC 1185
 |||||||
 Db 1180 TTGGCGGTGCTTTTGGCCCATATGAAAGTTTCTGACTTCAGAAAGCCTGTGGCATGC 1239
 QY 1186 CATGAATTTGCTGGGCTACCGCCACGTCAATCTGACACAGACAGAAACCTGTGCT 1245
 |||||||
 Db 1240 CATGAATTTGCTGGGCTACCGCCACGTCAATCTGACACAGACAGAAACCTGTGCT 1299
 QY 1246 GGATCTTCTGCTACCTGAGATGCTCCGAGCAATCCATCAAGAGGTTGAGAAGGCT 1305
 |||||||
 Db 1300 GGATCTTCTGCTACCTGAGATGCTCCGAGCAATCCATCAAGAGGTTGAGAAGGCT 1359
 QY 1306 TGTGCGACCTGTGTGTCAGCTCAGTCACTTCTCTGAAATGCTTCTGAGCCTGCTACA 1365
 |||||||
 Db 1360 TGTGCGACCTGTGTGTCAGCTCAGTCACTTCTCTGAAATGCTTCTGAGCCTGCTACA 1419
 QY 1366 TCTGTGACCTTAACTACATGCTGTGTGGATTCACACTGAGTGTGAGTGTGTCCACAG 1425
 |||||||
 Db 1420 TCTGTGACCTTAACTACATGCTGTGTGGATTCACACTGAGTGTGAGTGTGTCCACAG 1479
 QY 1426 TGTCTCAAGCANAAGAGCTTTGTGTCAATGCTTGTCTTAAAAAACAGACTGGGAACT 1485
 |||||||
 Db 1480 TGTCTCAAGCANAAGAGCTTTGTGTCAATGCTTGTCTTAAAAAACAGACTGGGAACT 1539
 QY 1486 TATGTGAGACACATCCACCAATGAAACAGGGTATTTGCTCTTCTTCTTGTGATCT 1545
 |||||||
 Db 1540 TATGTGAGACACATCCACCAATGAAACAGGGTATTTGCTCTTCTTCTTGTGATCT 1599
 QY 1546 TCTGTGTGGCAGACTTCAAGACTTTGTGGCCTGAGAGGCTTTAAGCAGCAGAGT 1605
 |||||||
 Db 1600 TCTGTGTGGCAGACTTCAAGACTTTGTGGCCTGAGAGGCTTTAAGCAGCAGAGT 1659
 QY 1606 ATCAGTGAATTTGATCCATAAATCCCTGTGTCACATCTTGCCAAATGGGGAATGATCT 1665
 |||||||
 Db 1660 ATCAGTGAATTTGATCCATAAATCCCTGTGTCACATCTTGCCAAATGGGGAATGATCT 1719
 QY 1666 TTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703

Db 1720 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1757

RESULT 3

ID	Accession	Standard	CDNA	Length (bp)
AA16947	standard	CDNA	1333	BP

AC AAS16947;
xy

DT	12-MAR-2002	(first entry)
YY		

Human L-selectin sulfotransferase-2 (LSST-2) cDNA

KM Human; Beta1,3ant; betaal.3-N-acetylgalactosaminyltransferase; MEOA-79; ss;
KM L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
KM ulcerative colitis; inflammatory skin disorder; psoriasis; Lichen planus;
KM allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
KM delayed-type hypersensitivity reaction; hyperplastic thymus; antilucer;
KM antiinflammatory; antipsoriatic; antidiabetic; dermatological;
KM antiallergic.

OS Homo sapiens.

EH	Key	Location/Qualifiers
EH	CDS	111 1253

FT	CDS	111..1253
----	-----	-----------

FT
yy
/product= "Human LSST-2"

PN WO200185177-A1.
YY

PD 15-NOV-2001
XX

PE 10-MAY-2001; 2001MO-US15452.
XX

PR 11-MAY-2000; 200005-0369320;
XX XX

(BORN -) BORNHALL INSIDE

XX
XX
XX

[illegible]

DR P-PSDB; AAU11274.

PT New enzyme, useful

PT New enzyme, useful for modifying acceptor molecule, comprises an isolated L-selectin sulfotransferase-2 that directs expression of L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, on intestinal GlcNAc 6-sulfotransferase

PS Claim 19; Fig 4; 98pp; English

The present invention provides a method of modifying an acceptor molecule by contacting the acceptor with an isolated

CC fragment, where beta1,3GnT directs expression of a MECA-79 antigen. The
CC invention also provides a method of treating or preventing an
CC infection, also provides a method of treating or preventing an
CC infection, also provides a method of treating or preventing an

invention also provides a method of treating or preventing an L-selectin-mediated condition by reducing the expression or activity of beta1-3ant that directs expression of a UICN-70 anti- α - β 1-3ant that directs expression of a UICN-70 anti- α - β 1-3ant.

that inhibits binding of I-selectin to a MECA-79 antigen for example by administering to the subject an oligosaccharide I-selectin antagonist by administering expression of a MECA-79 antigen. This can be done

administering antibody material that specifically binds beta₁3GNT, and/or a beta₁3GNT antisense nucleic acid molecule. 1-selectin and/or a beta₁3GNT antisense nucleic acid molecule, for example by

CC sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression

CC fragment can be reduced in combination with reducing the expression of
CC activity of beta1,3gnt. The method is useful for treating b-selectin

CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis,

CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC

XX Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;
SQ

Query Match	Similarity	61.8%	Score 1262:	DB 24:	Length 1333:
Best Local	Similarity	100.0%:	Pred. No. 0:		
Matches 1262:	Conservative	0:	Mismatches	0:	Indels
					Gaps 0:
QY	107	AAGGCTCTCCACTTCAGCACAATGCTACTGCGCTAAAAAAATGAAGCTCCTGCTGTTCTG	166		
Db	72	AAGGCTCTCCACTTCAGCACAATGCTACTGCGCTAAAAAAATGAAGCTCCTGCTGTTCTG	131		
QY	167	GTTTCCAGATGCGCATCTTGGCTCTATTCTTCCATGTACAGCACAATCATGCTCC	226		
Db	132	GTTTCCAGATGCGCATCTTGGCTCTATTCTTCCATGTACAGCACAATCATGCTCC	191		
QY	227	CTGTCTATGAAGGACACAGCCGAGCGGATGACGCGTGTCTGTCTCTCGGGCGCT	286		
Db	192	CTGTCTATGAAGGACACAGCCGAGCGGATGACGCGTGTCTGTCTCTCGGGCGCT	251		
QY	287	GCGCTCTCTTCTTGTGGGGGACGTTTTTGGGAGCACCACAGATGTTTTCTACCTGATGGAG	346		
Db	252	GCGCTCTCTTCTTGTGGGGGACGTTTTTGGGAGCACCACAGATGTTTTCTACCTGATGGAG	311		
QY	347	CCCCCGTGGCAGCGTGTGATACCTTTCAGAGAGACACCGCGTGATGCTGCACATGGCT	406		
Db	312	CCCCCGTGGCAGCGTGTGATACCTTTCAGAGAGACACCGCGTGATGCTGCACATGGCT	371		
QY	407	GTCGGGGATCTGATATACGGGCGGTCTTCTTGTGCGCATGAGCGCTTTGATGCTTACATG	466		
Db	372	GTCGGGGATCTGATATACGGGCGGTCTTCTTGTGCGCATGAGCGCTTTGATGCTTACATG	431		
QY	467	GAACCTGGTCCCGGAGACAGTCCAGCGCTTTTCAGTGGGAGAAACGCGGCGCCTGTGT	526		
Db	432	GAACCTGGTCCCGGAGAGACAGTCCAGCGCTTTTCAGTGGGAGAAACGCGGCGCCTGTGTGT	491		
QY	527	TCTCCACCTGGCTGTGACATCATCCCAATGAATGAATCCCGCGGGGCTCACATGACAG	586		
Db	492	TCTCCACCTGGCTGTGACATCATCCCAATGAATGAATCCCGCGGGCTCACATGACAG	551		
QY	587	CTCCCTGTCACATCAACAGCCCTTTGAGGTGGTGGAGAAAGCGCGCGCTCTACAGCCAC	646		
Db	552	CTCCCTGTCACATCAACAGCCCTTTGAGGTGGTGGAGAAAGCGCGCTCTCTACAGCCAC	611		
QY	647	GTGGTGTCTCAAGAGAGGTGCGCTTCTTCAACTGCAAGTCCCTCTACCGCGCTGCTAAAGAC	706		
Db	612	GTGGTGTCTCAAGAGAGGTGCGCTTCTTCAACTGCAAGTCCCTCTCTGAAAGAC	671		
QY	707	CCCTCCCTCAACCTGCATATGTGTGACCTGTGTCCGGAGACCCCGGCGCGGTGTCCGTTCC	766		
Db	672	CCCTCCCTCAACCTGCATATGTGTGACCTGTGTCCGGAGACCCCGGCGCGGTGTCCGTTCC	731		
QY	767	CGAGAACGCACAAAGGAGATCTCATGATTTAGACAGTCGATGTGATGGGCGACAGCATAG	826		
Db	732	CGAGAACGCACAAAGGAGATCTCATGATTTAGACAGTCGATGTGATGGGCGACAGCATAG	791		
QY	827	CAAAAACCTCAAGAAAGAGGAGACCAACCTCACTATGTGATGACAGGTCTATGCGCAAAAGCAG	886		
Db	792	CAAAAACCTCAAGAAAGAGGAGACCAACCTCACTATGTGATGACAGGTCTATGCGCAAAAGCAG	851		
QY	887	CTGGAGATCTTCAAGAACCATTCAGTCCCTGTGCGCAAGGCGCGAGGAAGCGTACTGCTT	946		
Db	852	CTGGAGATCTTCAAGAACCATTCAGTCCCTGTGCGCAAGGCGCGAGGAAGCGTACTGCTT	911		
QY	947	GTCGGCTATGAGGACCTGGCTGCGAGCCCTTGGCGCCAGACTTCCGATGTATGAAATTC	1006		
Db	912	GTCGGCTATGAGGACCTGGCTGCGAGCCCTTGGCGCCAGACTTCCGATGTATGAAATTC	971		
QY	1007	GTCGGGATTTGAATTTCTTGCCCATCTTTCAGACCTGGGTGTCATTAATCATCACCGAGGCAAG	1066		
Db	972	GTCGGGATTTGAATTTCTTGCCCATCTTTCAGACCTGGGTGTCATTAATCATCACCGAGGCAAG	1031		
QY	1067	GGCATGGGTGACACACGCTTTTCCACACAATAATGCAAGGATGGCCCTTAATGTCTCCAGGCT	1126		
Db	1032	GGCATGGGTGACACACGCTTTTCCACACAATAATGCAAGGATGGCCCTTAATGTCTCCAGGCT	1091		


```

Db      991 TTGTGGCTATGAGGAGCTGGCTGAGAGCCCTGTGGCCAGACTCCCGATATATGAAT 1050
QY      1005 TCGTGGGATTTGGAATCTTGCCCATCTTCAGACTGGGTGATATACATACCCGAGCA 1064
Db      1051 TGTGGGATTTGGAATCTTGCCCATCTTCAGACTGGGTGATATACATACCCGAGCA 1110
QY      1065 AGGGCATGGGTGACACGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGG 1124
Db      1111 AGGGCATGGGTGACACGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGG 1170
QY      1125 CTGGGCTGTGCTTGGCCCTATGAAAGGTTCTGAGCTCGAAAGCCTGTGGGATG 1184
Db      1171 CTGGGCTGTGCTTGGCCCTATGAAAGGTTCTGAGCTCGAAAGCCTGTGGGATG 1230
QY      1185 CCATGAATTTGCTGGGCTACCGCCAGCTGATCTGACAAAGAACAGAAACCTGTGC 1244
Db      1231 CCATGAATTTGCTGGGCTACCGCCAGCTGATCTGACAAAGAACAGAAACCTGTGC 1290
QY      1245 TGGATTTGCTGTCTACTGAGACTGCTCCCTGAGCAATCCACTAAGAGGTTGAAGGCT 1304
Db      1291 TGGATTTGCTGTCTACTGAGACTGCTCCCTGAGCAATCCACTAAGAGGTTGAAGGCT 1350
QY      1305 TTGCTGCCACCTGCTGACGCTCAGTCACTTCTCTGAAATGCTTCTGAGCCTTGCTAC 1364
Db      1351 TTGCTGCCACCTGCTGACGCTCAGTCACTTCTCTGAAATGCTTCTGAGCCTTGCTAC 1410
QY      1365 ATCTCTGAGCCTTAACATGCTGTGGGTATCACACTGATGATGATGTTGTCCACAC 1424
Db      1411 ATCTCTGAGCCTTAACATGCTGTGGGTATCACACTGATGATGATGTTGTCCACAC 1470
QY      1425 GTCCTGAGAGAGAGACTTTGTGTCATGCTGTGTCTAGAAACAGACCTGGGAAC 1484
Db      1471 GTCCTGAGAGAGAGACTTTGTGTCATGCTGTGTCTAGAAACAGACCTGGGAAC 1530
QY      1485 TTATGTGAGCAGACATCCACCAGTGAACAGAGGATTTGCTTCTTCTTCTTGATC 1544
Db      1531 TTATGTGAGCAGACATCCACCAGTGAACAGAGGATTTGCTTCTTCTTCTTGATC 1590
QY      1545 TTCCCTGCTGGGAGACTTCAGAGACTTTGTGGCTGGAGGCTTATTAGCAGACACAG 1604
Db      1591 TTCCCTGCTGGGAGACTTCAGAGACTTTGTGGCTGGAGGCTTATTAGCAGACACAG 1650
QY      1605 TTACAGTGAATTTGATCATTAACCTCCCTGTCACATCTTGCCCAATGGGGAATGATC 1664
Db      1651 TTACAGTGAATTTGATCATTAACCTCCCTGTCACATCTTGCCCAATGGGGAATGATC 1710
QY      1665 TTTCACCAAAAGAGCTCAGCAGCATTTTCCACAGAGATGC 1703
Db      1711 TTTCACCAAAAGAGCTCAGCAGCATTTTCCACAGAGATGC 1749

RESULT 5
ID      AAK91803 standard; cDNA; 877 BP.
XX      AAK91803;
XX      06-NOV-2001 (first entry)
XX      Human cDNA 5'-end sequence, SEQ ID NO: 263.
XX      Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX      Homo sapiens.
XX      EP130094-A2.
XX      05-SEP-2001.
XX      07-JUL-2000; 2000EP-0114089.
XX      08-JUL-1999; 99JP-0194486.

```

```

PR      11-JAN-2000; 2000JP-0118774.
PR      02-MAY-2000; 2000JP-0183765.
XX      (HELI-) HELIX RES INST.
PI      Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI      Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX      WPI; 2001-524255/58.
PT      830 Primers useful for synthesizing full length cDNA clones and their
PT      use in genetic manipulation -
PS      Claim 2; SEQ ID NO 263; 1380bp + sequence listing; English.
XX      The invention relates to primers for synthesizing full length cDNA
XX      clones. 830 cDNA molecules encoding a human protein have been
XX      isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
XX      molecules have been determined. Primers for synthesizing the full length
XX      cDNA are useful for clarifying the function of the protein encoded by
XX      the cDNA. The full length clones were obtained by construction of full
XX      length enriched cDNA libraries that were synthesised by the oligo-capping
XX      method. The primers enable the production of the full length cDNA easily
XX      without any special methods. The present sequence is the nucleotide
XX      sequence of the 5'-end of a cDNA provided in the invention.
XX      Note: The sequence data for this patent did not form part of the printed
XX      specification, but was obtained in CD-ROM format directly from EPO.
SQ      Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;
XX
Query Match      34.0%; Score 695; DB 22; Length 877;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      106 CAAGTCTTCCACTTACAGCACAATGCTACTGCTTAATAAATAGCTGCTGTTCT 165
Db      47 CAAGTCTTCCACTTACAGCACAATGCTACTGCTTAATAAATAGCTGCTGTTCT 106
QY      166 GGTTCCTCCAGATGGCCATCTGGCTATTTCTTCACATGTACACCAACATCAGCTC 225
Db      107 GGTTCCTCCAGATGGCCATCTGGCTATTTCTTCACATGTACACCAACATCAGCTC 166
QY      226 CCTGTATGAAGCAGACGCCGAGCGCATGACAGCTGCTGCTGCTCCGAGGCTC 285
Db      167 CCGTCTATGAAGCAGACGCCGAGCGCATGACAGCTGCTGCTGCTCCGAGGCTC 226
QY      286 TGGCTCTTCTTTGTGGGCGAGCTTTTGGGCGACGCCAGATGTTTCTACCTGATGA 345
Db      227 TGGCTCTTCTTTGTGGGCGAGCTTTTGGGCGAGGCCAGATGTTTCTACCTGATGA 286
QY      346 GCCCGCTGGCAGCTGTGATGTACCTTCAAGCAGACACCGCTGGATGTGACATGGC 405
Db      287 GCCCGCTGGCAGCTGTGATGTACCTTCAAGCAGACACCGCTGGATGTGACATGGC 346
QY      406 TGTGGGATCTGATACGGGCGCTCTTCTGTGCGACATGAGGCTCTTGATGCTACAT 465
Db      347 TGTGGGATCTGATACGGGCGCTCTTCTGTGCGACATGAGGCTCTTGATGCTACAT 406
QY      466 GGAACCTGTGCTCCCGAGAGACAGTCCAGCTCTTTCAGTGGGAGAACAGCGCGGCTGTG 525
Db      407 GGAACCTGTGCTCCCGAGAGACAGTCCAGCTCTTTCAGTGGGAGAACAGCGCGGCTGTG 466
QY      526 TTCTGACACTGCTGTGACATCATCTCCACAAAGTAAATCATCCCGGGGCTCAGCAG 585
Db      467 TTCTGACACTGCTGTGACATCATCTCCACAAAGTAAATCATCCCGGGGCTCAGCAG 526
QY      586 GCTCCTGTGAGTCAACAGACGCCCTTGTAGGATGGGAGAGGCTGCGCTCCACAGCCA 645
Db      527 GCTCCTGTGAGTCAACAGACGCCCTTGTAGGATGGGAGAGGCTGCGCTCCACAGCCA 586
QY      646 CGTGATGCTCAAGAGAGTGCAGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAGA 705
Db      587 CGTGATGCTCAAGAGAGTGCAGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAGA 646

```


QY 706 CCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGGGCGGTTCCTTC 765
 DB 647 CCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGGGCGGTTCCTTC 706
 QY 766 CCGAGAACGACAAAGGAGATCTCATGATTGACA 800
 DB 707 CCGAGAACGACAAAGGAGATCTCATGATTGACA 741

RESULT 6
 ID AAK93921 standard: cDNA: 877 BP.
 AC AAK93921;
 XX
 DE 06-NOV-2001 (first entry)
 DT
 XX Human cDNA clone representative sequence, SEQ ID NO: 2381.
 DE Human cDNA clone representative sequence, SEQ ID NO: 2381.
 KW Human; full length cDNA: cDNA synthesis; oligo-capping; ss.
 XX Homo sapiens.
 OS
 XX EPI130094-A2.
 PN
 XX 05-SEP-2001.
 PD
 XX 07-JUL-2000; 2000EP-0114089.
 PE
 XX 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX WPI; 2001-524255/58.
 DR
 XX 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Example 11: SEQ ID NO 2381; 1380bp + sequence listing: English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesized by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence was used in
 CC representative sequence from a human clone which was used in
 CC homology searches to identify the clone.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 SQ Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;

Query Match 34.0%; Score 695; DB 22; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 CAAAGTCTTCCACTTACGACACATGCTACTGCTTAAATAAATGAAGCTCTGCTGTTCT 165
 DB 47 CAAAGTCTTCCACTTACGACACATGCTACTGCTTAAATAAATGAAGCTCTGCTGTTCT 106
 QY 166 GGTTCGCCAGATGGCATCTGGCTCTATCTTCTCCACATGTACAGCCACACATCAGCTC 225
 DB 107 GGTTCGCCAGATGGCATCTGGCTCTATCTTCTCCACATGTACAGCCACACATCAGCTC 166

QY 226 CTTGCTATGAAGGACAGACCCGAGCGATGACATGCTGCTGCTCTTCCTGCGCTC 285
 DB 167 CTTGCTATGAAGGACAGACCCGAGCGATGACATGCTGCTGCTCTTCCTGCGCTC 226
 QY 286 TGGCTCTTCTTTTGGGGGAGCTTTTGGGGAGCACCAGATGTTTCTACATGATGGA 345
 DB 227 TGGCTCTTCTTTTGGGGGAGCTTTTGGGGAGCACCAGATGTTTCTACATGATGGA 286
 QY 346 GCCCGCTTGGCAGCTGTGATGACCTTCAAGCAGACGCCCTGATGCTGACATGGC 405
 DB 287 GCCCGCTTGGCAGCTGTGATGACCTTCAAGCAGACGCCCTGATGCTGACATGGC 346
 QY 406 TGTGCGGATCTGATACGAGGCGCTTCTGTCGACATGAGCGCTTGTGATGCTTACAT 465
 DB 347 TGTGCGGATCTGATACGAGGCGCTTCTGTCGACATGAGCGCTTGTGATGCTTACAT 406
 QY 466 GGAACCTGTGTCGCCGAGACAGTCCAGCCTTTTCAGTGGGAGAACACCGGCGCTGTG 525
 DB 407 GGAACCTGTGTCGCCGAGACAGTCCAGCCTTTTCAGTGGGAGAACACCGGCGCTGTG 466
 QY 526 TTTCTGACCTGCTGCTGACATCATCCCAAGATGAATATCCCGGCGCTCACTGAG 585
 DB 467 TTTCTGACCTGCTGCTGACATCATCCCAAGATGAATATCCCGGCGCTCACTGAG 526
 QY 586 GTCCTGTGCAATCAAGAGCCCTTTGAGTGGTGGAGAGGCGCTGCTCTACAGCA 645
 DB 527 GTCCTGTGCAATCAAGAGCCCTTTGAGTGGTGGAGAGGCGCTGCTCTACAGCA 586
 QY 646 GGTGTGCTCAAGAGAGTGGCTTTCTTCAACCTGAGTCCCTTACCGCTGCTGAAGA 705
 DB 587 GGTGTGCTCAAGAGAGTGGCTTTCTTCAACCTGAGTCCCTTACCGCTGCTGAAGA 646
 QY 706 CCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGGGCGGTTCCTTC 765
 DB 647 CCCCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGGGGCGGTTCCTTC 706
 QY 766 CCGAGAACGACAAAGGAGATCTCATGATTGACA 800
 DB 707 CCGAGAACGACAAAGGAGATCTCATGATTGACA 741

RESULT 7
 ID ABV89280 standard: cDNA: 505 BP.
 AC ABV89280;
 XX
 DE 13-DEC-2002 (first entry)
 DT
 XX Human colon cancer related cDNA SEQ ID NO 2595.
 DE Human colon cancer related cDNA SEQ ID NO 2595.
 KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
 XX ss.
 OS Homo sapiens.
 XX
 PN W0200258534-A2.
 PD 01-AUG-2002.
 PE 19-NOV-2001; 2001WO-US43704.
 PF 20-NOV-2000; 2000US-252222P.
 PR 06-FEB-2001; 2001US-267011P.
 PR 28-MAR-2001; 2001US-279670P.
 PR 10-JUL-2001; 2001US-304037P.
 XX (CORI-) CORIXA CORP.
 PI Stolk JA, Xu J, Chenault RA, Meagher MJ, Secrist H, King GE;
 XX WPI; 2002-608400/65.
 DR

XX New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer
 XX
 PS Claim 1; SEQ ID NO 2595; 266bp + Sequence Listing; English.
 CC
 CC The invention relates to a human colon tumour expressed polynucleotide
 CC (I) encoding a polypeptide (II), ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV6669-ABV8289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i) under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (ii) and (iii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at http://wipo.int/pub/published_pcl_sequences.
 CC
 SO Sequence 505 BP; 120 A; 125 C; 119 G; 141 T; 0 other;

Query Match 19.7%; Score 403; DB 24; Length 505;
 Best Local Similarity 100.0%; Pred. No. 2.6e-189;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCTTGGTCCACCTGCTGTCAGCTCAGTCACTTCTGTAATGCTTGTGAGCCTTGC 1360
 DB 1 GGCTTGGTCCACCTGCTGTCAGCTCAGTCACTTCTGTAATGCTTGTGAGCCTTGC 60
 QY 1361 CTACATCTCTGAGCCTTAACATACATCTCTGTGGGATACACAGTGTGAGTGTCTC 1420
 DB 61 CTACATCTCTGAGCCTTAACATACATCTCTGTGGGATACACAGTGTGAGTGTCTC 120
 QY 1421 AACGCGTCTCAAGCAGAGACTTGTGTCATGCTTGTGTCATGTAAGAAACACACCTGGG 1480
 DB 121 AACGCGTCTCAAGCAGAGACTTGTGTCATGCTTGTGTCATGTAAGAAACACACCTGGG 180
 QY 1481 AACCTTATGAGACACATCCACAGTGAAGAGGATATGCTTCTTCTTCTTCTT 1540
 DB 181 AACCTTATGAGACACATCCACAGTGAAGAGGATATGCTTCTTCTTCTTCTTCTT 240
 QY 1541 GATCTTCTGTCGTGGGACAGACTTGTGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1600
 DB 241 GATCTTCTGTCGTGGGACAGACTTGTGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 300
 QY 1601 ACAGTATCAGTGAATGATCCATAAAGCTCCCTGTCACATCTTGGCCCAATGGGAATG 1660
 DB 301 ACAGTATCAGTGAATGATCCATAAAGCTCCCTGTCACATCTTGGCCCAATGGGAATG 360
 QY 1661 GATCTTCAACAAGAGCTACACAGATTTTCCACAGAGATGC 1703
 DB 361 GATCTTCAACAAGAGCTACACAGATTTTCCACAGAGATGC 403

RESULT 8
 ABR54724
 ID ABR54724 standard; cDNA; 517 BP.
 XX
 AC ABR54724;
 XX
 AC
 XX 18-JUN-2002 (first entry)
 XX
 DE Human colon cancer-associated cDNA, SEQ ID NO 194.
 XX
 KM Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200212280-A2.
 XX
 PD 14-FEB-2002.

XX
 XX 30-JUL-2001; 2001WO-US23826.
 XX
 PF
 XX 03-AUG-2000; 2000US-223265P.
 PR
 XX 02-OCT-2000; 2000US-227406P.
 PR
 XX 20-MAR-2001; 2001US-277495P.
 PR
 XX 03-JUL-2001; 2001US-302702P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Pyle RA, Xu J, Secrist H;
 XX
 XX WPI; 2002-257462/30.
 DR
 XX
 XX Novel polynucleotide encoding colon tumour polypeptides, useful as
 PT vaccines for treating colon cancers
 PT
 XX
 PS Claim 1; Page 206; 425bp; English.

CC The invention relates to isolated polynucleotides (I) encoding colon
 CC tumour polypeptides (II). (I) is useful for stimulating an immune
 CC response in a patient and treating colon cancer in a patient.
 CC Oligonucleotides derived from (I) are useful for determining the presence
 CC of cancer in a patient. (I) and (II) are useful in pharmaceutical
 CC compositions, e.g. vaccines, and other compositions for the diagnosis
 CC and treatment of colon cancer. A composition comprising a first component
 CC selected from physiologically acceptable carriers and immunostimulants,
 CC and an antigen-presenting cell expressing (II) is useful for inhibiting
 CC development of cancer in a patient. (I) is useful in the design and
 CC preparation of ribozyme molecules for inhibiting expression of tumour
 CC polypeptides and (I). ABR54531-ABR54544 represent human colon cancer cDNA
 CC sequences of the invention.
 CC
 SO Sequence 517 BP; 121 A; 130 C; 124 G; 142 T; 0 other;

Query Match 19.7%; Score 403; DB 24; Length 517;
 Best Local Similarity 100.0%; Pred. No. 2.6e-189;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCTTGGTCCACCTGCTGTCAGCTCAGTCACTTCTGTAATGCTTGTGAGCCTTGC 1360
 DB 13 GGCTTGGTCCACCTGCTGTCAGCTCAGTCACTTCTGTAATGCTTGTGAGCCTTGC 72
 QY 1361 CTACATCTCTGAGCCTTAACATACATCTCTGTGGGATACACAGTGTGAGTGTCTC 1420
 DB 73 CTACATCTCTGAGCCTTAACATACATCTCTGTGGGATACACAGTGTGAGTGTCTC 132
 QY 1421 AACGCGTCTCAAGCAGAGACTTGTGTCATGCTTGTGTCATGTAAGAAACACACCTGGG 1480
 DB 133 AACGCGTCTCAAGCAGAGACTTGTGTCATGCTTGTGTCATGTAAGAAACACACCTGGG 192
 QY 1481 AACCTTATGAGACACATCCACAGTGAAGAGGATATGCTTCTTCTTCTTCTT 1540
 DB 193 AACCTTATGAGACACATCCACAGTGAAGAGGATATGCTTCTTCTTCTTCTTCTT 252
 QY 1541 GATCTTCTGTCGTGGGACAGACTTGTGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1600
 DB 253 GATCTTCTGTCGTGGGACAGACTTGTGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 312
 QY 1601 ACAGTATCAGTGAATGATCCATAAAGCTCCCTGTCACATCTTGGCCCAATGGGAATG 1660
 DB 313 ACAGTATCAGTGAATGATCCATAAAGCTCCCTGTCACATCTTGGCCCAATGGGAATG 372
 QY 1661 GATCTTCAACAAGAGCTACACAGATTTTCCACAGAGATGC 1703
 DB 373 GATCTTCAACAAGAGCTACACAGATTTTCCACAGAGATGC 415

RESULT 9
 AAC76156
 ID AAC76156 standard; cDNA; 2988 BP.
 XX
 AC AAC76156;

08-FEB-2001 (first entry)

Human OREFX ORF1711 polynucleotide sequence SEQ ID NO:3421.

Human; open reading frame; OREFX; detection; cytosolic; hepatotropic; vulnary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective; anticonvulsant; osteoprotic; antirheumatic; immunosuppressant; cardiant; immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antihypertensive; antiviral; antibacterial; antifungal; antineoplastic; antithyroid; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; cholesterol ester storage; systemic lupus erythematosus; infection; severe combined immunodeficiency; malaria; autoimmune disorder; asthma; allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound; bone damage; cartilage damage; antiinflammatory disease; coagulation; thrombosis; contraceptive; ss.

Homo sapiens.

MO200058473-A2.

05-OCT-2000.

31-MAR-2000; 2000MO-US08621.

31-MAR-1999; 99US-0127607.

02-APR-1999; 99US-0127636.

05-APR-1999; 99US-0127728.

30-MAR-2000; 2000US-0540763.

(CURA-) CURAGEN CORP.

Shinkets RA, Leach M;

WPI: 2000-602362/57.

P-PSDB: AAB41947.

Novel nucleic acids and peptides derived from open reading frame X, useful for treating e.g. cancers, proliferative disorders, neurodegenerative disorders and cardiovascular disease -

Claim 5; Page 2597-2599; 5507pp; English.

AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397, which represent the human OREFX open reading frames 1 to 3161. The OREFX sequences have activities such as: cytostatic; hepatotropic; vulnary; antiparasitic; antiparkinsonian; neurotropic; neuroprotective; osteoprotic; anticonvulsant; antirheumatic; immunosuppressant; immunostimulant; cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive; dermatological; immunosuppressive; antihypertensive; antibacterial; antiviral; antifungal; antineoplastic; antithyroid; antianemic; gene therapy; cancer; proliferative disorder; hypertension; neurodegenerative disorder; osteoarthritis; graft vs host disease; cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS; severe combined immunodeficiency (SCID); AIDS; viral, bacterial or fungal infection, malaria, autoimmune disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to enhance coagulation; to inhibit thrombosis; and as a contraceptive.

Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other;

Query Match 15.0%; Score 306; DB 21; Length 2988;

Best Local Similarity 100.0%; Pred. NO. 3.7e-141; Indels 0; Gaps 0;

Matches 306; Conservative 0; Mismatches 0;

QY	452	TTTGATGCTACACATGAGAACCGTGGTCCCGGAGACAGTCCAGCCCTTTCACTGGGAGAAC	511
Db	1480	TTTGATGCTACACATGAGAACCGTGGTCCCGGAGACAGTCCAGCCCTTTCACTGGGAGAAC	1538
QY	512	AGCGGGGCGCTGTGTTCTGCACTCGCCTGTGACATCATCCCAAGATGAATCATATCC	571
Db	1540	AGCGGGGCGCTGTGTTCTGCACTCGCCTGTGACATCATCCCAAGATGAATCATATCC	1599
QY	572	CGGCTACACGACAGGCTTCCTGTGACATCACAGCCCTTTGAGGTGTGTGACAAGGCTTGC	631
Db	1600	CGGCTACACGACAGGCTTCCTGTGACATCACAGCCCTTTGAGGTGTGTGACAAGGCTTGC	1659
QY	632	CGGCTACACGACAGGCTTCCTGTGACATCACAGGAGGTGGCTTCTTCAACCTGACATCCCTTAC	691
Db	1660	CGGCTACACGACAGGCTTCCTGTGACATCACAGGAGGTGGCTTCTTCAACCTGACATCCCTTAC	1719
QY	692	CCGCTGTGAAAGACCCCTCCCTCAACCTGCATATGTCGACCTGTCCGGACCCCGG	751
Db	1720	CCGCTGTGAAAGACCCCTCCCTCAACCTGCATATGTCGACCTGTCCGGACCCCGG	1779
QY	752	GCCGCTG 757	
Db	1780	GCCGCTG 1785	
RESULT 10			
ID	ABK54794	standard; cDNA; 389 BP.	
XX	ABK54794;		
AC	ABK54794;		
XX	18-JUN-2002 (first entry)		
DT	18-JUN-2002 (first entry)		
XX	Human colon cancer-associated cDNA, SEQ ID No 264.		
DE	Human colon cancer; immunogenic; vaccine; tumour; gene; ss.		
KW	Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.		
XX	Homo sapiens.		
OS	WO200212280-A2.		
PN	WO200212280-A2.		
XX	14-FEB-2002.		
PD	14-FEB-2002.		
XX	30-JUL-2001; 2001WO-US23826.		
PF	03-AUG-2000; 2000US-223265P.		
XX	02-OCT-2000; 2000US-2237406P.		
PR	20-MAR-2001; 2001US-277495P.		
PR	03-JUL-2001; 2001US-302702P.		
XX	(CORI-) CORIXA CORP.		
PA	Pyle RA, Xu J, Secretist H;		
XX	WPI: 2002-257462/30.		
PI	Novel polynucleotide encoding colon tumour polypeptides, useful as		
DR	vaccines for treating colon cancers		
XX	Claim 1; Page 225; 425pp; English.		
PS	The invention relates to isolated polynucleotides (I) encoding colon		
XX	tumour polypeptides (II). (I) is useful for stimulating an immune		
CC	response in a patient and treating colon cancer in a patient.		
CC	Oligonucleotides derived from (I) are useful for determining the presence		
CC	of cancer in a patient. (I) and (II) are useful in pharmaceutical		
CC	compositions, e.g. vaccines, and other compositions for the diagnosis		
CC	and treatment of colon cancer. A composition comprising a first component		
CC	selected from physiologically acceptable carriers and immunostimulants,		
CC	and an antigen-presenting cell expressing (II) is useful for inhibiting		
CC	development of cancer in a patient. (I) is useful in the design and		
CC	preparation of ribozyme molecules for inhibiting expression of tumour		

CC polypeptides and (I). ABK54531-ABK5464 represent human colon cancer cDNA
 CC sequences of the invention.
 XX
 SQ Sequence 389 BP; 121 A; 83 C; 76 G; 109 T; 0 other;
 Query Match 5.6%; Score 114; DB 24; Length 389;
 Best Local Similarity 100.0%; Pred. No. 6.9e-46;
 Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1930 TGGCAGCTGTTAACTAAATTCACAAATAGGTTCTGTTAGAAATGCCCTTTTATG 1989
 DB 266 TGGCAGCTGTTAACTAAATTCACAAATAGGTTCTGTTAGAAATGCCCTTTTATG 325
 QY 1990 CTTCCTTAATTAATAGCAGTAATGTCATTTTATGGATCCTAACAAAAA 2043
 DB 326 CTTCCTTAATTAATAGCAGTAATGTCATTTTATGGATCCTAACAAAAA 379
 RESULT 11
 AAK92588/c
 ID AAK92588 standard; cDNA; 548 BP.
 AC AAK92588;
 XX
 DT 06-NOV-2001 (first entry)
 DE Human cDNA 3'-end sequence, SEQ ID NO: 1048.
 XX
 KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP1130094-A2.
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 PI WPI; 2001-524255/58.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Claim 3; SEQ ID NO 1048; 1380BP + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is the nucleotide
 CC sequence of the 3'-end of a cDNA provided in the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 SQ Sequence 548 BP; 160 A; 109 C; 129 G; 148 T; 2 other;
 Query Match 5.1%; Score 105; DB 22; Length 548;
 Best Local Similarity 100.0%; Pred. No. 2e-41;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGGCAGCTGTTAACTAAATTCACAAATAGGTTCTGTTAGAAATGCCCTTTTATG 1989
 DB 108 TGGCAGCTGTTAACTAAATTCACAAATAGGTTCTGTTAGAAATGCCCTTTTATG 49
 QY 1990 CTTCCTTAATTAATAGCAGTAATGTCATTTTATGGATCCCTAA 2034
 DB 48 CTTCCTTAATTAATAGCAGTAATGTCATTTTATGGATCCCTAA 4
 RESULT 12
 AAD24670
 ID AAD24670 standard; cDNA; 1647 BP.
 AC AAD24670;
 XX
 DT 12-MAR-2002 (first entry)
 DE Human drug metabolizing enzyme (DME)-5 cDNA.
 XX
 KW Human; drug metabolizing enzyme; gene therapy; autoimmune disorder;
 KW inflammatory disorder; acquired immune deficiency syndrome; infection;
 KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
 KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
 KW cancer; endocrine disorder; hypothyroidism; pituitary disorder;
 KW gastrointestinal disorder; metabolic disorder; developmental disorder;
 KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
 KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
 KW DME-5; ss.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1188
 FT /*tag- a
 FT /product- "Human DME-5 protein"
 FT sig-peptide 1..96
 FT /*tag- b
 FT mat-peptide 97..1185
 FT /*tag- c
 FT /product- "Human mature DME-5 protein #1"
 FT sig-peptide 1..105
 FT /*tag- d
 FT mat-peptide 106..1185
 FT /*tag- e
 FT /product- "Human mature DME-5 protein #2"
 FT MO200179468-A2.
 XX
 PD 25-OCT-2001.
 XX
 PF 12-APR-2001; 2001MO-US11869.
 XX
 PR 13-APR-2000; 2000US-197590P.
 PR 19-APR-2000; 2000US-198403P.
 PR 28-APR-2000; 2000US-200185P.
 PR 05-MAY-2000; 2000US-202234P.
 PR 11-MAY-2000; 2000US-203509P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Policky JL, Hafalla A, Burford N, Ring HZ, Lal P, Tribouley CM;
 PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;
 PI Gandhi AK, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;
 PI An-Young J;
 PI WPI; 2002-066363/09.
 DR P-PDB; AAE15438.
 XX
 PT Novel isolated human drug metabolizing enzymes referred as DME 1-10,
 PT useful for diagnosing, treating, or preventing disorders associated
 PT with aberrant expression of DME such as allergy, anemia, asthma,
 PT infertility -
 XX

PS Claim 5; Page 139; 143pp; English.

XX The invention relates to human drug metabolizing enzymes referred as
 CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides
 CC of the invention are useful for assessing toxicity of test compounds
 CC and in gene therapy. Sequences of the invention are useful in the
 CC diagnosis, prevention and treatment of autoimmune/inflammatory
 CC disorders such as acquired immune deficiency syndrome (AIDS), adult
 CC respiratory distress syndrome, allergies, anaemia, atherosclerosis,
 CC asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's
 CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's
 CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple
 CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus
 CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,
 CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic
 CC infections; cell proliferative disorders such as actinic keratosis,
 CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,
 CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,
 CC epilepsy; endocrine disorders such as disorders of the hypothalamus
 CC and pituitary resulting from lesions such as primary brain tumours,
 CC adenomas, infarction associated with pregnancy, aneurysms, vascular
 CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,
 CC glaucoma, pigmentosa; metabolic disorders such as Addison's disease,
 CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,
 CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,
 CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such
 CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,
 CC hyperillirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
 CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
 CC disorders. The present sequence is human DME-5 protein cDNA.

XX Sequence 1647 BP; 287 A; 560 C; 501 G; 299 T; 0 other;

XX

Query Match 2.9%; Score 59; DB 24; Length 1647;
 Best Local Similarity 100.0%; Pred. No. 1.4e-18;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 624 AGGCGTCCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCCTTTCACCTGCAG 682
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 488 AGGCGTCCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCCTTTCACCTGCAG 546

RESULT 13
 AAD02700 standard; cDNA; 1694 BP.

XX
 AC AAD02700;
 XX
 DT 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-4beta (GST-4beta) cDNA.

XX Human: glycosyl sulfotransferase-4beta; GST-4beta; Immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocardiitis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;
 KW chromosome 16q23.1; ss.

XX
 OS Homo sapiens.

XX
 XX Key Location/Qualifiers
 FH 5'UTR 9..188
 FT /tag- a
 FT 189..1376
 FT CDS /tag- b
 FT /product- "Human glycosyl sulfotransferase-4beta
 FT (GST-4beta)"
 FT /note- "CDS is referred as SEQ ID NO:21 in brief

FT description of the figures (page no: 4)"
 FT 3'UTR 1377..1694
 FT /tag- c

XX WO200106015-A1.
 XX 25-JAN-2001.
 XX 19-JUL-2000; 2000WO-US19741.
 XX 20-JUL-1999; 99US-0144694.
 XX 13-JUL-2000; 2000US-0593828.
 XX (REGC) UNIV CALIFORNIA.
 XX Rosen SD, Lee JK, Hemmerlich S;
 XX WPI; 2001-138471/14.
 XX P-PSDB; AAY72640.

PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications

XX Claim 6; Fig 4A; 128pp; English.

XX The present sequence is human glycosyl sulfotransferase-4beta (GST-4
 CC beta) cDNA. GST-4 gene is found on chromosome 16q23.1.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adrenailitis, hypoparathyroidism, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocardiitis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.

XX Sequence 1694 BP; 259 A; 586 C; 514 G; 335 T; 0 other;

XX

Query Match 2.9%; Score 59; DB 22; Length 1694;
 Best Local Similarity 100.0%; Pred. No. 1.4e-18;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 624 AGGCGTCCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCCTTTCACCTGCAG 682
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 676 AGGCGTCCGCTCTACAGCCAGCGTGCTCAAGAGAGTGCCTTTCACCTGCAG 734

RESULT 14
 AAD02699 standard; cDNA; 2044 BP.

XX
 AC AAD02699;
 XX
 DT 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.

XX Human: glycosyl sulfotransferase-4alpha; GST-4alpha; Immunosuppressive;
 KW therapy; selectin binding inhibitor; gene therapy; inflammation;
 KW systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 KW polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 KW glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;
 KW Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
 KW demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 KW myocardiitis; adult respiratory distress syndrome; eczema; psoriasis;
 KW asthma; hypersensitivity; rheumatic fever; tissue rejection;

• Fri Aug 15 08:09:27 2003

us-09-645-078-1.ing

Page 13

Db 834 AGGCTGCGGCTCTACAGCCACGTGGTGCCTCAAGAGAGTGGCTTCTTCACCTGCAG 892

Search completed: August 15, 2003, 04:50:47
Job time : 545 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 15, 2003, 04:13:39 ; Search time 3887 Seconds
(Without alignments)
12774.388 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043
Sequence: 1 gaattccattgtgttgta.....tgggacccaataaaaaaa 2043

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 1215238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 38370230

Minimum DB seq length: 300

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estin:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inu:*
19: em_gss_pln:*
20: em_gss_vrl:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_pbg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	685	33.5	954	12	BI823850 603039012
2	579	28.3	583	2	BK490456 Homo sapi
3	534	26.1	668	9	AL709927 DKF2P8860
4	479	23.4	571	12	BM129080 1f17c04.y

Result No.	Score	Query Match	Length	ID	Description
5	322	15.8	593	12	BM969292 UI-CF-ENO
6	290	14.2	660	10	BE878439 MRO-ET010
7	288	14.1	680	13	Bu684395 UT-CF-ENO
8	218	10.7	553	12	BM128370
9	218	10.7	553	12	BM128370
10	107	5.2	362	12	BM128370
11	105	5.1	417	9	AM002418
12	101	4.9	358	10	BM056840
13	74	3.6	419	9	BM056840
14	59	2.9	525	10	BM197521
15	59	2.9	620	9	BM197521
16	59	2.9	695	10	BM197521
17	55	2.9	735	10	BM197521
18	55	2.3	721	28	BM197521
19	47	2.3	436	9	BM197521
20	27	1.3	507	12	BM245312
21	27	1.3	571	12	BM245312
22	27	1.3	695	12	BM245312
23	27	1.3	783	12	BM245312
24	27	1.3	852	12	BM245312
25	27	1.3	852	12	BM245312
26	27	1.3	1923	11	BM245312
27	27	1.1	362	10	BM245312
28	27	1.1	397	9	BM245312
29	27	1.1	687	14	BM245312
30	27	1.1	692	14	BM245312
31	27	1.1	834	13	BM245312
32	27	1.1	361	9	BM245312
33	27	1.1	493	10	BM245312
34	27	1.1	501	28	BM245312
35	27	1.1	516	9	BM245312
36	27	1.1	527	4	BM245312
37	27	1.1	910	13	BM245312
38	27	1.1	1881	10	BM245312
39	27	1.0	340	29	BM245312
40	27	1.0	410	28	BM245312
41	27	1.0	424	29	BM245312
42	27	1.0	510	14	BM245312
43	27	1.0	541	28	BM245312
44	27	1.0	616	28	BM245312
45	27	1.0	631	28	BM245312

ALIGNMENTS

RESULT 1
BI823850
LOCUS
DEFINITION
603039012F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179826 5',
mRNA sequence.
ACCESSION
BI823850
VERSION
BI823850.1 GI:15935400
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
NIH-MGC http://mgc.ncl.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapds-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: LLM11448 row: h column: 03
High quality sequence stop: 856.

QY 281 CGCTGTGCTCTCTCTTTTGTGGGACGCTTTTGGGACGACCCAGATGTTTCTACTG 340
 |||||
 Db 241 CGCTGTGCTCTCTCTTTTGTGGGACGCTTTTGGGACGACCCAGATGTTTCTACTG 300
 |||||
 QY 341 ATGAGACCCGCGTGGACGCTGTGGATGACCTTCAAGACAGACACGCGCTGGATGCTGAC 400
 |||||
 Db 301 ATGAGACCCGCGTGGACGCTGTGGATGACCTTCAAGACAGACACGCGCTGGATGCTGAC 360
 |||||
 QY 401 ATGAGCTTGGGGATCTGATACGAGCGCGCTCTCTTGTGGACATGACGCTTGTGATGCC 460
 |||||
 Db 361 ATGAGCTTGGGGATCTGATACGAGCGCGCTCTCTTGTGGACATGACGCTTGTGATGCC 420
 |||||
 QY 461 TACATGAACTGTGCTCCCGAGACAGTCCAGCCTTTTCAAGTGGAGAACAGCGCGGACC 520
 |||||
 Db 421 TACATGAACTGTGCTCCCGAGACAGTCCAGCCTTTTCAAGTGGAGAACAGCGCGGACC 480
 |||||
 QY 521 CTGTGTCTGACACCTGCTGTGACATGATCCACAGATGAATATCATCCCGGCGTCAAC 580
 |||||
 Db 481 CTGTGTCTGACACCTGCTGTGACATGATCCACAGATGAATATCATCCCGGCGTCAAC 540
 |||||
 QY 581 TGCAGGCTCTGTGACATCAACAGACCCCTTGGAGTGTG 619
 |||||
 Db 541 TGCAGGCTCTGTGACATCAACAGACCCCTTGGAGTGTG 579
 |||||

RESULT 3

AL709927 668 bp mRNA linear EST 12-JUN-2003
 LOCUS DKEZp8602364_r1 666 (synonym: hlcc3) Homo sapiens cDNA clone

ACCESSION AL709927
 AL709927.1 GI:19693282

VERSION EST.
 KEYWORDS Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 668)

AUTHORS Bahr, A., Lauber, J., Mewes, H.W., Well, B., Aml, C., Oeanger, A., Fob
 EST (Bahr, A., Lauber, J., Mewes, H.W., Well, B., et al.)
 G., Han, M. and Wiemann, S.
 Unpublished
 CONTACT: Bahr A

TITLE
 JOURNAL
 COMMENT

MPIS

Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert.
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 sequenced by Qlagen (Hilden/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 No sl sequence available.

This clone (DKEZp8602364) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 Location/Qualifiers

FEATURES

source

1..668
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKEZp8602364"
 /tissue_type="human skeletal muscle"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_id="666 (synonym: hlcc3)"
 /note="vector: pTRIPLEX2; Site_1: SfilI; Site_2: SfilI;
 cDNA-collection"

BASE COUNT 132 a 202 c 178 g 154 t 2 others
 ORIGIN

Query Match 26.1%; Score 534; DB 9; Length 668;
 Best Local Similarity 100.0%; Pred. NO. 2.3e-209;
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 AGAGAGAAAAAGCGATGCGCCGCTAGCAGTGAAGCTCTCAAAAGACAGAGGAGCC 99
 |||||
 Db 6 AGAGAGAAAAAGCGATGCGCCGCTAGCAGTGAAGCTCTCAAAAGACAGAGGAGCC 65
 |||||
 QY 100 AAGCCAAAGGCTCTTCCACTTACAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCT 159
 |||||
 Db 66 AAGCCAAAGGCTCTTCCACTTACAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCT 125
 |||||
 QY 160 GTTCTGTGTTCCAGATGCGCATCTTGCTCTTCTTCTCACTATGACAGCAACAT 219
 |||||
 Db 126 GTTCTGTGTTCCAGATGCGCATCTTGCTCTTCTTCTCACTATGACAGCAACAT 185
 |||||
 QY 220 CAGTCCCTGTCTTGAAGGACAGCCCGAGCGATGACGCTGGTGTCTTCTGCTG 279
 |||||
 Db 186 CAGTCCCTGTCTTGAAGGACAGCCCGAGCGATGACGCTGGTGTCTTCTGCTG 245
 |||||
 QY 280 GCGCTGTGCTCTTCTTGTGGGACGCTTTTGGGAGCAGCCAGATGTTTCTACT 339
 |||||
 Db 246 GCGCTGTGCTCTTCTTGTGGGACGCTTTTGGGAGCAGCCAGATGTTTCTACT 305
 |||||
 QY 340 GATGAGACCCGCGTGGACAGTGTGATGACCTTCAAGCAGACACCGCTGGATGCTGCA 399
 |||||
 Db 306 GATGAGACCCGCGTGGACAGTGTGATGACCTTCAAGCAGACACCGCTGGATGCTGCA 365
 |||||
 QY 400 CATGCTGTGCGGATCTGATGACGAGCGCGCTCTTGTGGACATGAGCGCTTGTATGC 459
 |||||
 Db 366 CATGCTGTGCGGATCTGATGACGAGCGCGCTCTTGTGGACATGAGCGCTTGTATGC 425
 |||||
 QY 460 CTACATGGAACCTGTGCTCCCGAGACAGTCCAGCCTTTTCAAGTGGAGAAACAGCGGCG 519
 |||||
 Db 426 CTACATGGAACCTGTGCTCCCGAGACAGTCCAGCCTTTTCAAGTGGAGAAACAGCGGCG 485
 |||||
 QY 520 CCTGTGTCTGACACCTGCTGTGACATGATCCCAAGATGAATATCATCCCGG 573
 |||||
 Db 486 CCTGTGTCTGACACCTGCTGTGACATGATCCCAAGATGAATATCATCCCGG 539
 |||||

RESULT 4

BM129080

LOCUS BM129080 571 bp mRNA linear EST 12-MAR-2002
 DEFINITION I17C04.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 cDNA clone IMAGE:5676655 5' similar to TR:09YR3 09YR3
 N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE. ; mRNA sequence.

ACCESSION BM129080
 BM129080.1 GI:17123632

VERSION EST.
 KEYWORDS Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 571)

AUTHORS Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
 Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, A.,
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas
 M., Gibbons, M., McCann, R., Cole, R., Tsagarelashvili, R., Williams, T.,
 Jackson, Y. and Bowers, Y.

Unpublished
 CONTACT: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Other ESTs: I17C04.x1
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138

Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@dioph.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Juliana Brown
 (brown@fas.harvard.edu) This sequence now available from the IMAGE
 consortium, for clone orders contact: info@image.llnl.gov

FEATURES High quality sequence stop: 434.
Location/Qualifiers
source
1. 571
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5676655"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Mellon Normalized Human Islet 4 NA-HIS 1"
/note="Organ: Pancreas; Vector: pSPORT1; Site:1: Not 1;
Site:2: Sal 1; Starting library constructed using
SuperScript Plasmid Library Kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an Ecot of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxapatite chromatography and used to make this
library."

BASE COUNT 109 a 169 c 143 g 150 t

ORIGIN

Query Match 23.4%; Score 479; DB 12; Length 571;
Best Local Similarity 100.0%; Pred. No. 1.2e-186;
Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

106 CAGGCTTCACCTTCAGCACAATGCTACTGCTTAAATAAGTACCTCTGCTGTTCT 165
111
111
73 CAGGCTTCACCTTCAGCACAATGCTACTGCTTAAATAAGTACCTCTGCTGTTCT 132
111
111
166 GGTTCGCCAGATGGCCATCTTGCTCTATCTTCCACATGTACAGCCACAATCAGCTC 225
111
111
133 GGTTCGCCAGATGGCCATCTTGCTCTATCTTCCACATGTACAGCCACAATCAGCTC 192
111
111
226 CCGTCTATGAAGGACACAGCCGAGGAGCAGTGTGCTGCTTCCGCGGCCTC 285
111
111
193 CCGTCTATGAAGGACACAGCCGAGGAGCAGTGTGCTGCTTCCGCGGCCTC 252
111
111
286 TGGCTCTCTTTTGTGGGACAGCTTTTGGGACAGCAGATGTTTCTACCTGATGGA 345
111
111
253 TGGCTCTCTTTTGTGGGACAGCTTTTGGGACAGCAGATGTTTCTACCTGATGGA 312
111
111
346 GCGCGCGTGACAGTGTGATGACCTTCAAGCAGACACCCCTGATGCTGACATGGC 405
111
111
313 GCGCGCGTGACAGTGTGATGACCTTCAAGCAGACACCCCTGATGCTGACATGGC 372
111
111
406 TGTGGGAGATGATACGAGGCGCTTCTTGTGCGACATGAGGCTTTGATGCTACAT 465
111
111
373 TGTGGGAGATGATACGAGGCGCTTCTTGTGCGACATGAGGCTTTGATGCTACAT 432
111
111
466 GGAACCTGTGCTCCCGGAGACAGTCCAGCCTTTTCAGTGGGAGAACGCCGCGCTGTG 525
111
111
433 GGAACCTGTGCTCCCGGAGACAGTCCAGCCTTTTCAGTGGGAGAACGCCGCGCTGTG 492
111
111
526 TTCTGCACTGCTGTGACATCATCCACAGATGAATCATCCCGGCGCTACTGCA 584
111
111
493 TTCTGCACTGCTGTGACATCATCCACAGATGAATCATCCCGGCGCTACTGCA 551
111
111

RESULT 5
BM969292/c 593 bp mRNA linear EST 20-FEB-2003
LOCUS
DEFINITION UI-CF-ENO-acp-1-21-0-UI.s1 UI-CF-ENO Homo sapiens cDNA clone
ACCESSION BM969292
VERSION BM969292.1 GI:19586879

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 593)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
CDNA Library preparation: Dr. M. Bento Soares, University of Iowa
CDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES Location/Qualifiers
source
1. 593
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-ENO-acp-1-21-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-ENO"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site:1: EcoR I; Site:2: Not I;
UI-CF-ENO is a cDNA library containing the following
tissue(s): Primary lung Cystic Fibrosis Epithelial Cells.
The library was constructed according to Bonaldo, Lennon
and Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dr)18 tail. The sequence tag for this
library is CTGCTCAGGT.
TAG_LIB=UI-CF-ENO
TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG_SEQ=CTGCTCAGGT"

BASE COUNT 157 a 127 c 137 g 172 t

ORIGIN

Query Match 15.8%; Score 322; DB 12; Length 593;
Best Local Similarity 100.0%; Pred. No. 6.6e-122;
Matches 322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1382 ACATGCTGTGGGATACACAGAGTGTGAGTGTGTCACAGGCTCAGACAGAGA 1441
111
111
593 ACATGCTGTGGGATACACAGAGTGTGAGTGTGTCACAGGCTCAGACAGAGA 534
111
111
1442 CTTTGTGTCATGCTGTGTAGAAAACACAGTGGGACCTTATGTGACACACAT 1501
111
111
533 CTTTGTGTCATGCTGTGTAGAAAACACAGTGGGACCTTATGTGACACACAT 474
111
111
1502 CCACACAGTGAACAGGGTATTGCTCTCTTTCTTGATCTCTGCTGCGCAGAC 1561
111
111

```

Db      473  CCCACGAGTGAACAGGGTATGTGCTCTTCTTTTGTGATCTCTGCTGCGAGAC 414
QY      1562  TTTCAGACATTTTGGGCTGGAGCCCTATTAGACACACAGATATAGTGAATGATC 1621
Db      413  TTTCAGACATTTTGGGCTGGAGCCCTATTAGACACACAGATATAGTGAATGATC 354
QY      1622  CATTAACCTCCCTGCTCCACATCTTGCCCAATGGGAAATGATCTTTCACCAAGAGCTCA 1681
Db      353  CATTAACCTCCCTGCTCCACATCTTGCCCAATGGGAAATGATCTTTCACCAAGAGCTCA 294
QY      1682  CCAGCATTTTCACACAGATGC 1703
Db      293  CCAGCATTTTCACACAGATGC 272

RESULT 6
LOCUS   BF878439                680 bp    mRNA    linear    EST 17-JAN-2001
DEFINITION  MR0-ET0109-191100-002-h06 ET0109 Homo sapiens cDNA, mRNA sequence.
ACCESSION  BF878439
VERSION    BF878439.1  GI:12268569
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 680)
AUTHORS   Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
          Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
          Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
          Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
          ,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
          Simpson,A.J.
          Shotgun sequencing of the human transcriptome with ORF expressed
          sequence tags
          Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
          10737800
COMMENT    Contact: Simpson A.J.G.
          Laboratory of Cancer Genetics
          Ludwig Institute for Cancer Research
          Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
          Brazil
          Tel: +55-11-2704922
          Fax: +55-11-2707001
          Email: asimpson@ludwig.org.br
          This sequence was derived from the FAPESP/LICR Human Cancer Genome
          Project. This entry can be seen in the following URL
          (http://www.ludwig.org.br/scripts/gethtml2.pl?tl=MR06t2-MR0-ET0109-
          191100-002-h06&t3=2000-11-19&t4=1)
          Seq primer: puc 18 forward
          High quality sequence start: 17
          High quality sequence stop: 641.
          Location/Qualifiers
            1..680
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /dev_stage="Adult"
              /clone_id="ET0109"
              /note="Organ: lung, tumor; Vector: puc18; Site_1: Smal;
              Site_2: SmaI; A mini-library was made by cloning products
              derived from ORESTES PCR (U.S. Letters Patent application
              No. 196,716 - Ludwig Institute for Cancer Research)
              profiles into the pUC 18 vector. Reverse transcription of
              tissue mRNA and cDNA amplification were performed under
              low stringency conditions."

BASE COUNT  168 a 164 c 163 g 184 t 1 others
ORIGIN
Query Match 14.2%; Score 290; DB 10; Length 680;
Best Local Similarity 100.0%; Pred. No. 9,9e-109;

```

```

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY      1414  TGTGTCCACACGCTGCTCAAGAGAAAGACTTTTGTTCATGCTTGTGTCTAGAAAACAG 1473
Db      106  TGTGTCCACACGCTGCTCAAGAGAAAGACTTTTGTTCATGCTTGTGTCTAGAAAACAG 165
QY      1474  ACTGGGAACTTATGTAGACACATCCACCAGTGAAGAAAGAGATATTCCTTCTC 1533
Db      166  ACTGGGAACTTATGTAGACACATCCACCAGTGAAGAAAGAGATATTCCTTCTC 225
QY      1534  TTTTCTTGATCTTCCTGCTGCGAGACTTCAGACTTGTGGCTGGAGCCCTATTAA 1593
Db      226  TTTTCTTGATCTTCCTGCTGCGAGACTTCAGACTTGTGGCTGGAGCCCTATTAA 285
QY      1594  GCACGACACAGTATCAGTGAATGATTCATTAACCTCCCTGCTCCATCTTGCCCAATG 1653
Db      286  GCACGACACAGTATCAGTGAATGATTCATTAACCTCCCTGCTCCATCTTGCCCAATG 345
QY      1654  GGGATGAGATCTTCACCAAGAGCTCACAGCATTTTTCACACAGATGC 1703
Db      346  GGGATGAGATCTTCACCAAGAGCTCACAGCATTTTTCACACAGATGC 395

RESULT 7
LOCUS   BU684395/c                669 bp    mRNA    linear    EST 07-OCT-2002
DEFINITION  UI-CF-ENO-aco-f-08-0-UI-s1 UI-CF-ENO Homo sapiens cDNA clone
ACCESSION  BU684395
VERSION    BU684395.1  GI:23537302
KEYWORDS   EST.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE  1 (bases 1 to 669)
AUTHORS   Bernaldo,M.F., Lennon,G. and Soares,M.B.
          Normalization and subtraction: two approaches to facilitate gene
          discovery
          Genome Res. 6 (9), 791-806 (1996)
          97044477
COMMENT    Contact: McCray, PB
          McCray Lab
          University of Iowa
          2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
          Tel: 319 356 4866
          Fax: 319 356 7171
          Email: paul-mccray@uiowa.edu
          Tissue Procurement: Dr. M. J. Welsh, University of Iowa
          CDNA library preparation: Dr. M. Bento Soares, University of Iowa
          CDNA library arrayed by: Dr. M. Bento Soares, University of Iowa
          DNA sequencing by: Dr. M. Bento Soares, University of Iowa
          Clone Distribution: Researchers may obtain clones from Research
          Genetics (www.resgen.com) or from Open Biosystems
          (www.openbiosystems.com).
          Seq primer: M13 FORWARD
          POLY-A=yes
          Location/Qualifiers
            1..669
              /organism="Homo sapiens"
              /mol_type="mRNA"
              /db_xref="taxon:9606"
              /clone_id="UI-CF-ENO-aco-f-08-0-UI"
              /tissue_type="Primary Lung Cystic Fibrosis Epithelial
              Cells"
              /dev_stage="Adult"
              /lab_host="DH10B (Life Technologies) (T1 phage resistant)"
              /clone_id="UI-CF-ENO"
              /note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
              modified PolyLinker; Site_1: EcoR I; Site_2: Not I;
              UI-CF-ENO is a cDNA library containing the following
              tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells."

```

The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)₁₈ tail. The sequence tag for this library is CTGCTCAGGT.

TAG: TISSUE=Human Lung Epithelial Cell Lines untreated LPS
TAG_LTB=UI-CF-ENO
6hr to LPS 24h

BASE COUNT 191 a 138 c 147 g 193 t
ORIGIN TAG_SRO-CTGCTCAGGT

Query Match 14.1%; Score 288; DB 13; Length 669;
Best Local Similarity 100.0%; Pred. No. 6.7e-108;
Matches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1416 TGTCCACAGCTGCTCAGCAGAGAGACTTTGTGTCATGCTGTGTAGAAACAGAC 1475
DB 636 TGTCCACAGCTGCTCAGCAGAGAGACTTTGTGTCATGCTGTGTAGAAACAGAC 577
QY 1476 TGGGGAACCTTATGTAGACAGCAGATCCACAGAGGATTTGCTCTTCTT 1535
DB 576 TGGGGAACCTTATGTAGACAGCAGATCCACAGAGGATTTGCTCTTCTT 517
QY 1536 TTCTTGATCTTCCTGCTGCGAGACTTCAGACATTTGGCTGAGGCCATTAAAG 1595
DB 516 TTCTTGATCTTCCTGCTGCGAGACTTCAGACATTTGGCTGAGGCCATTAAAG 457
QY 1596 AGCAGACAGTATCAGTGAATTCATTAACCTCCGTCCACATCTTGCCCAATGGG 1655
DB 456 AGCAGACAGTATCAGTGAATTCATTAACCTCCGTCCACATCTTGCCCAATGGG 397
QY 1656 GAATGATCTTCACCAAGAGCTCCAGCATTTCCACAGAGATGC 1703
DB 396 GAATGATCTTCACCAAGAGCTCCAGCATTTCCACAGAGATGC 349

RESULT 8
BM128370/c 553 bp mRNA linear EST 12-MAR-2002
LOCUS lft1d12.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens

DEFINITION cDNA clone IMAGE:567635 3', mRNA sequence.
ACCESSION BM128370
VERSION BM128370.1 GI:17122922

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 553)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, R., Lemisha, I., Pearce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Riter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarelshvili, R., Williams, T., Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium
Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138

Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@molb.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center for information on

obtaining a clone please contact: Juliana Brown
(brownj@molb.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 425.

FEATURES

location/Qualifiers
1.553
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_image="IMAGE:567635"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/note="Organ: Pancreas; Vector: pSPORT1; Site: 1; Not 1; Site: 2; Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoR of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 163 a 111 c 127 g 152 t
ORIGIN

Query Match 10.7%; Score 218; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.3e-79;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1486 TATGTAGACAGCAGATCCACAGGATGATGCTCTCTCTTTTGTATCT 1545
DB 553 TATGTAGACAGCAGATCCACAGGATGATGCTCTCTCTTTTGTATCT 494
QY 1546 TTCTGCTGCGAGACTTCAGAGACTTTGGCTGAGGACATTAAAGCAGACAG 1605
DB 493 TTCTGCTGCGAGACTTCAGAGACTTTGGCTGAGGACATTAAAGCAGACAG 434
QY 1606 ATCAGTGAATTCATTAACCTCCGTCCACATCTTGCCCAATGGGAATGATCT 1665
DB 433 ATCAGTGAATTCATTAACCTCCGTCCACATCTTGCCCAATGGGAATGATCT 374
QY 1666 TTACCAAGAGCTCCAGCATTTCCACAGAGATGC 1703
DB 373 TTACCAAGAGCTCCAGCATTTCCACAGAGATGC 336

RESULT 9
BM128831/c 553 bp mRNA linear EST 12-MAR-2002

LOCUS lft1c04.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
DEFINITION cDNA clone IMAGE:567655 3', mRNA sequence.
ACCESSION BM128831
VERSION BM128831.1 GI:17123383

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 553)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, R., Lemisha, I., Pearce, M., Brestelli, J., Gradwohl, G., Clifton, S., Hillier, L., Marra, M., Pape, D., Wylie, T., Martin, J., Blistain, A., Schmitt, A., Theising, B., Riter, E., Ronko, I., Bennett, J., Cardenas, M., Gibbons, M., McCann, R., Cole, R., Tsagarelshvili, R., Williams, T., Jackson, Y. and Bowers, Y.
Endocrine Pancreas Consortium

JOURNAL COMMENT

Unpublished

Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue

Endocrine Pancreas Consortium

Harvard University, Howard Hughes Medical Institute

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge, MA 02138

Tel: 617-495-1812

Fax: 617-495-8557

Email: dmelton@molb.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by: Washington University Genome Sequencing Center for information on obtaining a clone please contact: Juliana Brown (brownj@fas.harvard.edu) This sequence now available from the IMAGE Consortium, for clone orders contact: info@image.llnl.gov

High quality sequence stop: 425.

FEATURES

SOURCE

1. 553

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:567655"

/sex="Both"

/tissue_type="Islets of Langerhans"

/dev_stage="Adult"

/lab_host="DH10B"

/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"

/note="Organ: Pancreas; Vector: pSPOR1; Site: 1; Not 1; Site 2: Sal 1; Starting library constructed using Superscript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 163 a 111 c 127 g 152 t

ORIGIN

Query Match 10.7%; Score 218; DB 12; Length 553;

Best Local Similarity 100.0%; Pred. No. 5.3e-79;

Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1486 TATGTGAGCAGACATCCACAGAGGATTTGCTCTTCTTTTGATCT 1545

|||||

553 TATGTGAGCAGACATCCACAGAGGATTTGCTCTTCTTTTGATCT 494

1546 TCCGTCTGGGAGAGCTTCAGAGACTTGGGCGCTATTAAGCAGACAGT 1605

|||||

493 TCCGTCTGGGAGAGCTTCAGAGACTTGGGCGCTATTAAGCAGACAGT 434

1606 ATCAGTGAATTCATTAACCTCCCTGTCACATTTGCCAATGGGAATGATCT 1665

|||||

433 ATCAGTGAATTCATTAACCTCCCTGTCACATTTGCCAATGGGAATGATCT 374

1666 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1703

|||||

373 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 336

RESULT 10

AM002418/c 362 bp mRNA linear EST 27-OCT-1999

LOCUS wu1903.x1 NCI_CGAP_G6C Homo sapiens cDNA clone IMAGE:2524564 3',

DEFINITION mRNA sequence.

ACCESSION AM002418

VERSION AM002418.1 GI:5849334

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 362)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL COMMENT

Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/BLND at: www.bld.llnl.gov/bdrp/image/image.html

Seq primer: -400P from G1bco.

FEATURES

SOURCE

1. 362

Location/Qualifiers

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2524564"

/tissue_type="pooled germ cell tumors"

/lab_host="DH10B"

/clone_lib="NCI_CGAP_G6C"

/note="Vector: p7T3D-Pac (Pharmacia) with a modified polylinker; Site: 1; Not 1; Site 2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP_G6C was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clones 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 108 a 69 c 77 g 108 t

ORIGIN

Query Match 5.2%; Score 107; DB 9; Length 362;

Best Local Similarity 100.0%; Pred. No. 3.7e-33;

Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1930 TCCCATCTCTTAATCTAAATTTCCCAATAGCTTCTGTAGAAATCCCTTTTATG 1989

|||||

107 TCCCATCTCTTAATCTAAATTTCCCAATAGCTTCTGTAGAAATCCCTTTTATG 48

1990 CTCTTAATTTATACATAATGTTCAATTTTATGAGATCCTAATA 2036

|||||

47 CTCTTAATTTATACATAATGTTCAATTTTATGAGATCCTAATA 1

RESULT 11

AM572510/c 417 bp mRNA linear EST 13-MAR-2000

LOCUS xq18g11.x2 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2751044 3',

DEFINITION mRNA sequence.

ACCESSION AM572510

VERSION AM572510.1 GI:7237243

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 417)

AUTHORS NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP), Tumor Gene Index

JOURNAL COMMENT Unpublished

Contact: Robert Strausberg, Ph.D.

Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov/image/html/resources.shtml
 Seq primer: -40UP from Gibco
 High quality sequence stop: 416.
 Location/Qualifiers

FEATURES

source

1. 417
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2751044"
 /tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"
 /lab_host="DH10B"
 /clone_lib="NCI-CGAP_UCL"
 /note="Organ: uterus; Vector: pCMV-SPORT6; Site:1; Salt; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"

BASE COUNT 140 a 65 c 80 g 132 t
 ORIGIN
 Query Match 5.1%; Score 105; DB 9; Length 417;
 Best Local Similarity 100.0%; Pred. No. 2.3e-32;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGGCAGCTGTATTAATAATCCCAATAGAGTCTGTGATGTCCTTTTATG 1989
 Db 266 TGGCAGCTGTATTAATAATCCCAATAGAGTCTGTGATGTCCTTTTATG 207
 QY 1990 CTCTTAATATTAGCAGTAATGTCATTTTATGGAGCTCTAA 2034
 Db 206 CTCTTAATATTAGCAGTAATGTCATTTTATGGAGCTCTAA 162

RESULT 12 358 bp mRNA linear EST 16-OCT-2000
 BF056840 7k10612.x1 NCI-CGAP-GC6 Homo sapiens CDNA clone IMAGE:3443783 3',
 LOCUS mRNA sequence.
 DEFINITION
 ACCESSION BF056840 GI:10810736
 VERSION
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 358)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL, send email to: infoimage.llnl.gov
 Seq primer: -40UP from Gibco.

FEATURES

source

1. 358
 /organism="Homo sapiens"
 /mol_type="mRNA"

/db_xref="taxon:9606"
 /clone="IMAGE:3443783"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /clone_lib="NCI-CGAP-GC6"
 /note="Vector: pTR73D-Pac (Pharmacia) with a modified polylinker; Site:1: Not I; Site:2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP-GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 107 a 68 c 82 g 101 t
 ORIGIN
 Query Match 4.9%; Score 101; DB 10; Length 358;
 Best Local Similarity 100.0%; Pred. No. 1.1e-30;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1932 CCATCTGTATTAATAATCCCAATAGAGTCTGTGATGTCCTTTTATGCT 1991
 Db 101 CCATCTGTATTAATAATCCCAATAGAGTCTGTGATGTCCTTTTATGCT 42
 QY 1992 TCTTAATATTAGCAGTAATGTCATTTTATGGAGCTCT 2032
 Db 41 TCTTAATATTAGCAGTAATGTCATTTTATGGAGCTCT 1

RESULT 13 419 bp mRNA linear EST 13-MAR-2000
 AM572390/x2 NCI-CGAP_COL4 Homo sapiens CDNA clone IMAGE:2799634 3',
 LOCUS mRNA sequence.
 DEFINITION
 ACCESSION AM572390 GI:7237123
 VERSION
 KEYWORDS EST.

SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 419)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 CDNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: image.llnl.gov/image/html/resources.shtml
 Seq primer: -40UP from Gibco
 High quality sequence stop: 343.
 Location/Qualifiers

FEATURES

source

1. 419
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2799634"
 /tissue_type="moderately-differentiated adenocarcinoma"
 /lab_host="DH10B"
 /clone_lib="NCI-CGAP_COL4"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site:1: SalI; Site:2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"

BASE COUNT 122 a 78 c 89 g 130 t

ORIGIN

Query Match 3.6%; Score 74; DB 9; Length 419;
 Best Local Similarity 100.0%; Pred. No. 1.4e-19;
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1630 TCCCTGTCACATCTTCCCAATGGGAGATGATCTTTCACCAAGAGCTCACCAGCAT 1689
 DB 419 TCCCTGTCACATCTTCCCAATGGGAGATGATCTTTCACCAAGAGCTCACCAGCAT 360
 QY 1690 TTCACAGAGATGC 1703
 DB 359 TTCACAGAGATGC 346

RESULT 14
 BFL97521/c 525 bp mRNA linear EST 03-NOV-2000
 LOCUS 7684608.x1 NCI-CGAP_Kid11 Homo sapiens CDNA clone IMAGE:3642903 3'
 DEFINITION similar to TR:09UBY3 09UBY3 N-ACETYLGLUCOSAMINE
 6-O-SULFOTRANSFERASE. ; mRNA sequence.

ACCESSION BFL97521 GI:11086670
 VERSION BFL97521
 KEYWORDS EST.
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 525)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL, send email to:
 info@image.llnl.gov
 High quality sequence stop: 451.

FEATURES

source
 1. 525
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3642903"
 /lab_host="DH10B"
 /clone_id="NCI-CGAP_Kid11"
 /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
 a modified polylinker; Site 1: Not I; Site 2: Eco RI;
 Plasmid DNA from the normalized library NCI-CGAP_Kid3 was
 prepared, and ss circles were made in vitro. Following HRP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The driver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library
 (clones 1322376-1323911, 1456007-1456775, and
 1500552-1502855). Subtraction by Bento Soares and M.
 Fatima Bonaldo."
 BASE COUNT 90 a 163 c 187 g 85 t
 ORIGIN

Query Match 2.9%; Score 59; DB 10; Length 525;
 Best Local Similarity 100.0%; Pred. No. 2e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTCCGCTCTTACAGCAGCTGTGCTCAAGAGAGTGCGCTTCAACTGCAG 682
 DB 260 AGGCTCCGCTCTTACAGCAGCTGTGCTCAAGAGAGTGCGCTTCAACTGCAG 202

RESULT 15
 A1824100/c 620 bp mRNA linear EST 21-DEC-1999
 LOCUS w746c01.x1 NCI-CGAP_Lu19 Homo sapiens CDNA clone IMAGE:2405856 3'
 DEFINITION similar to TR:075667 075667 D7J11.6.4; contains PIRs.b2 PIR5
 repetitive element ; mRNA sequence.

ACCESSION A1824100 GI:5444771
 VERSION A1824100
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 620)
 AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index

JOURNAL Unpublished
 COMMENT Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www-bio.llnl.gov/dbp/image/image.html
 Insert Length: 917 Std Error: 0:00
 Seq primer: -40UP from GAPco
 High quality sequence stop: 490.

FEATURES

source
 1. 620
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2405856"
 /tissue_type="squamous cell carcinoma, poorly
 differentiated (4 pooled tumors, including primary and
 metastatic)."
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /clone_id="NCI-CGAP_Lu19"
 /note="Organ: lung; Vector: pT7T3D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 pooled lung tumor tissue, and was then primed with a Not I
 - oligo(dT) primer. Double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT7T3 vector. Library went through one round of
 normalization. Library constructed by Bento Soares and M.
 Fatima Bonaldo."
 BASE COUNT 106 a 195 c 227 g 89 t 3 others
 ORIGIN

Query Match 2.9%; Score 59; DB 9; Length 620;
 Best Local Similarity 100.0%; Pred. No. 1.9e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTCCGCTCTTACAGCAGCTGTGCTCAAGAGAGTGCGCTTCAACTGCAG 682
 DB 314 AGGCTCCGCTCTTACAGCAGCTGTGCTCAAGAGAGTGCGCTTCAACTGCAG 256

Search completed: August 15, 2003, 07:57:24
 Job time : 3892 secs

THIS PAGE BLANK (insert)

460 TGTGGGGATCTGATACGGGCGTCTTCTTGTCGACATGAGCGCTTTGATGCCTACAT 519

28	17	0.8	673	4	US-09-280-116-93	Sequence 93, Appl 1
29	17	0.8	813	4	US-09-252-991A-1600	Sequence 160, Ap
30	17	0.8	857	4	US-09-171-209-4	Sequence 4, Appl1
31	17	0.8	1358	4	US-09-328-352-1988	Sequence 8, Appl
32	17	0.8	1466	3	US-09-130-242-8	Sequence 81, Appl1
33	17	0.8	1524	4	US-09-252-991A-1651	Sequence 1651, Ap
34	17	0.8	1540	1	US-08-286-872-5	Sequence 5, Appl1
35	17	0.8	1572	4	US-09-612-964-1	Sequence 1, Appl1
36	17	0.8	1575	1	US-07-988-260B-2	Sequence 2, Appl1
37	17	0.8	1683	4	US-09-252-991A-11226	Sequence 11226, A
38	17	0.8	1764	1	US-08-036-355B-150	Sequence 150, App
39	17	0.8	1764	1	US-08-465-569-150	Sequence 150, App
40	17	0.8	1764	1	US-08-249-322A-150	Sequence 150, App
41	17	0.8	1764	1	US-08-469-526A-150	Sequence 150, App
42	17	0.8	1764	2	US-08-734-591A-150	Sequence 150, App
43	17	0.8	1764	2	US-08-469-660-150	Sequence 150, App
44	17	0.8	1764	3	US-08-341-018-57	Sequence 57, Appl
45	17	0.8	1764	3	US-08-470-335-150	Sequence 150, App

ALIGNMENTS

RESULT 1

US-09-045-284A-1

Sequence 1, Application US/09045284A

Patent No. 6265192

GENERAL INFORMATION:

APPLICANT: Bistrup, Annette

APPLICANT: Rosen, Steven D.

APPLICANT: Hemmerich, Stefan

TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3

FILE REFERENCE: 6510-107U01

CURRENT APPLICATION NUMBER: US/09/045,284A

CURRENT FILING DATE: 1998-03-20

NUMBER OF SEQ ID NOS: 9

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 1

LENGTH: 2032

TYPE: DNA

ORGANISM: Homo sapiens

US-09-045-284A-1

Query Match

Best Local Similarity 75.7%; Score 1547; DB 3; Length 2032;

Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0

106

CAAGGTCCTCCACTTACAGCACAATAGCTACTCCCTAAAAAATGAAGCTCTGCTGTTTCT

165

160

CAAGGTCCTCCACTTACAGCACAATAGCTACTCCCTAAAAAATGAAGCTCTGCTGTTTCT

211

166

GATTTCACAGATGGCCATCTTGCGCTATTCTTCCACTGATACGACCAACAACATACGCTC

222

226

CCCTCATATGAAGCACAAGCCCGAGCGCATCAGCTGCTGTTGCTTCTCTGCGGCTC

289

280

CTGTCTATGAGGACAGCCCGAGCGCATCAGCTGCTGTTGCTTCTCTGCGGCTC

333

286

TGGCTCTTCTTTGTGGGGAGCTTTTGGGACAGCACCACCAGATGTTTCTACCTGATGA

345

340

TGGCTCTTCTTTGTGGGGAGCTTTTGGGACAGCACCACCAGATGTTTCTACCTGATGA

399

346

GCCCGCTGGGACAGCTGTGTGATGACCTTCAACACAAGCAGCGCTGTGATGTGCACATGGC

407

400

GCCCGCTGGGACAGCTGTGTGATGACCTTCAACACAAGCAGCGCTGTGATGTGCACATGGC

455

406

TGTCTCGGATGTGATACGGCGCTCTTCTTGTGAGATGATGAGCTCTTTGATGTGCTACAT

465

460

TGTCTCGGATGTGATACGGCGCTCTTCTTGTGAGATGATGAGCTCTTTGATGTGCTACAT

519

466

GGAACCTGTGCTCCCGGAGACAGTCCAGCTCTTTTCACTGGGAGAACAGCGGCGCTGTG

525

Db 520 GGAACCTGGTCCCCGAGAGAGTCAGACCTCTTTCAGTGGGAGAAAGCCGGGCGCTGTG 579
QY 526 TTCTGCACTGCGCTGTGACATCATCCACAAAGATGAATCATCCCCGGGCTCACTGAG 585
Db 580 TTCTGCACTGCGCTGTGACATCATCCACAAAGATGAATCATCCCCGGGCTCACTGAG 639
QY 586 GCTCCGTGTGAGTCAACAGCCCTTTAGAGTGTGTGAGAAAGGCTGCCGCTCTACAGCCA 645
Db 640 GCTCCGTGTGAGTCAACAGCCCTTTAGAGTGTGTGAGAAAGGCTGCCGCTCTACAGCCA 699
QY 646 CGTGGTGTCAAGAGAGGTGGGCTTTCTCAACCTGACAGTCCCTTACCCGCTGTGAAAGA 705
Db 700 CGTGGTGTCAAGAGAGGTGGGCTTTCTCAACCTGACAGTCCCTTACCCGCTGTGAAAGA 759
QY 706 CCCCTCCCTCAACCTGATATGCTGACCTGTGTCCGGAGCCCGGGGCGCTGTTCGTTTC 765
Db 760 CCCCTCCCTCAACCTGATATGCTGACCTGTGTCCGGAGCCCGGGGCGCTGTTCGTTTC 819
QY 766 CCGAGAGACGACAAAGAGAGATCTCATATGATTGACAGTCCGATTTGTATGGGAGCAGATGA 825
Db 820 CCGAGAGACGACAAAGAGAGATCTCATATGATTGACAGTCCGATTTGTATGGGAGCAGATGA 879
QY 826 GCAAAAACCTCAAGAGAGAGAGACCAACCTACTATGTATGTAGTATGCTGCAAGACCA 885
Db 880 GCAAAAACCTCAAGAGAGAGAGACCAACCTACTATGTATGTAGTATGCTGCAAGACCA 939
QY 886 GCTGAGATCTCAAGAGACATCCAGTCTGTGCGCAAGCCCTGACAGAACGCTACCTGCT 945
Db 940 GCTGAGATCTCAAGAGACATCCAGTCTGTGCGCAAGCCCTGACAGAACGCTACCTGCT 999
QY 946 TGTGGCTTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCCGAATGTATGAAT 1005
Db 1000 TGTGGCTTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCCGAATGTATGAAT 1059
QY 1006 CGTGGATTTGAATTTCTGCCCATCTTTCAGACTGGGTGATATGATCATCCGAGGACAA 1065
Db 1060 CGTGGATTTGAATTTCTGCCCATCTTTCAGACTGGGTGATATGATCATCCGAGGACAA 1119
QY 1066 GGGCATGAGTGAACACGCTTTCCACAAATGCGAGGAGTCCCTTAATGTCTCCAGGC 1125
Db 1120 GGGCATGAGTGAACACGCTTTCCACAAATGCGAGGAGTCCCTTAATGTCTCCAGGC 1179
QY 1126 TTGGGCTGTGCTTTGCCCATGAAAAAGTTCTGACATTCAGAAAGCCGTGGCAGTGC 1185
Db 1180 TTGGGCTGTGCTTTGCCCATGAAAAAGTTCTGACATTCAGAAAGCCGTGGCAGTGC 1239
QY 1186 CATGAATTTGCTGGCTACCGCCACGCTGAGATCTGAACAAAGAACAGAAACCTGTGCT 1245
Db 1240 CATGAATTTGCTGGCTACCGCCACGCTGAGATCTGAACAAAGAACAGAAACCTGTGCT 1299
QY 1246 GGAATCTTGTCTACTGAGCTGTCCCTGAGCAATTCACATAAGAGGTTGAGAGGCTT 1305
Db 1300 GGAATCTTGTCTACTGAGCTGTCCCTGAGCAATTCACATAAGAGGTTGAGAGGCTT 1359
QY 1306 TGTGTCACCTGGTGTACAGCTCAATCTTCTGAAATGCTTGAAGCTTGGCTTACA 1365
Db 1360 TGTGTCACCTGGTGTACAGCTCAATCTTCTGAAATGCTTGAAGCTTGGCTTACA 1419
QY 1366 TGTGTCAGCCTTAACATACATGCTGTGGGTATCACAAGTGTGAGTTGTGTCAACAG 1425
Db 1420 TGTGTCAGCCTTAACATACATGCTGTGGGTATCACAAGTGTGAGTTGTGTCAACAG 1479
QY 1426 TGTGTCAGCAGAAAGACTTTTGTGTCAATGCTTGTGTCTAGAAACAGACTGGGAACT 1485
Db 1480 TGTGTCAGCAGAAAGACTTTTGTGTCAATGCTTGTGTCTAGAAACAGACTGGGAACT 1539
QY 1486 TATGTAGCAGACATCCACAGCAGTGAACAGGGTATGCTCTTCTCTTCTTCTATCT 1545
Db 1540 TATGTAGCAGACATCCACAGCAGTGAACAGGGTATGCTCTTCTCTTCTTCTATCT 1599
QY 1546 TCTGTCTGTGGAGACTTGTAGAGACTTGTGGCTGAGGCTTATTAAGCAGACAGT 1605
Db 1600 TCTGTCTGTGGAGACTTGTAGAGACTTGTGGCTGAGGCTTATTAAGCAGACAGT 1659

QY 1606 ATCAGTGAATTTGATCCATAAACCTCCCTGTCCACATCTTGGCCAAATGGGAATGATCT 1665
Db 1660 ATCAGTGAATTTGATCCATAAACCTCCCTGTCCACATCTTGGCCAAATGGGAATGATCT 1719
QY 1666 TTCACCAAGAGCTCACACAGATTTTCCACAGAGATGC 1703
Db 1720 TTCACCAAGAGCTCACACAGATTTTCCACAGAGATGC 1757

RESULT 2
US-09-190-911-2
; Sequence 2, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Blstrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; EARLIER FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-2

Query Match 75.7%; Score 1547; DB 4; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAAGTCTTTCACCTGACGACAAATGCTACTGCTTAAAAAATGAAGCTCTGTGTTCT 165
Db 160 CAAGTCTTTCACCTGACGACAAATGCTACTGCTTAAAAAATGAAGCTCTGTGTTCT 219
QY 166 GGTTCACAGATGGGCAATCTGCTATCTTCCACATGTACAGGACAAATCAAGCTC 225
Db 220 GGTTCACAGATGGGCAATCTGCTATCTTCCACATGTACAGGACAAATCAAGCTC 279
QY 226 CCGTCTATGAAGCAGACGCGGAGCGATGACGCTGTGCTTCTGCTGCGGCTC 285
Db 280 CCGTCTATGAAGCAGACGCGGAGCGATGACGCTGTGCTTCTGCTGCGGCTC 339
QY 286 TGGCTCTTCTTTTGTGGGCACTTTTGGGCAAGCACCAGATGTTTCTACTGATGA 345
Db 340 TGGCTCTTCTTTTGTGGGCACTTTTGGGCAAGCACCAGATGTTTCTACTGATGA 399
QY 346 GCCCGCTGGACAGTGTGATGACCTTCAAGCAGACAGCCGCTGTGATGCTGACATGG 405
Db 400 GCCCGCTGGACAGTGTGATGACCTTCAAGCAGACAGCCGCTGTGATGCTGACATGG 459
QY 406 TGTGCGGAGTGTATACGGGCGCTTCTTGTGCGACATGAGCGCTTTGATGCTACAT 465
Db 460 TGTGCGGAGTGTATACGGGCGCTTCTTGTGCGACATGAGCGCTTTGATGCTACAT 519
QY 466 GGAACCTGTGTCGGGAGACAGCTCTTTCAGTGGGAGAAAGCCGGGCGCTGTG 525
Db 520 GGAACCTGTGTCGGGAGACAGCTCTTTCAGTGGGAGAAAGCCGGGCGCTGTG 579
QY 526 TTCTGCACTGCGCTGTGACATCATCCACAAAGATGAATCATCCCCGGGCTCACTGAG 585
Db 580 TTCTGCACTGCGCTGTGACATCATCCACAAAGATGAATCATCCCCGGGCTCACTGAG 639
QY 586 GCTCTGTGTGAGTCAACAGCCCTTTAGAGTGTGTGAGAAAGGCTTGCCTCTACAGCA 645
Db 640 GCTCTGTGTGAGTCAACAGCCCTTTAGAGTGTGTGAGAAAGGCTTGCCTCTACAGCA 699

```

646 CGTGTCTCAGAGAGTGGCTTCTTCAACCTGACAGTCCCTGACCGCTGTGAAGA 705
700 CGTGTCTCAGAGAGTGGCTTCTTCAACCTGACAGTCCCTGACCGCTGTGAAGA 759
706 CCCCTCCCTAACCTGCATATCGTGCACCTGGTCCGGGACCCCGGGCCGTGTCGTTCC 765
760 CCCCTCCCTAACCTGCATATCGTGCACCTGGTCCGGGACCCCGGGCCGTGTCGTTCC 819
766 CCGGAGACGCAAAAGGAGATCTCATGATGACAGTCCGATGTGTATGGGGGAGCATGA 825
820 CCGGAGAGGCAAAAGGAGATCTCATGATGACAGTCCGATGTGTATGGGGGAGCATGA 879
826 GCAAAACTCAAGAGAGGAGACCAACCTACTATGTATGATGATGATGATGATGATGATG 885
880 GCAAAACTCAAGAGAGGAGACCAACCTACTATGTATGATGATGATGATGATGATGATG 939
886 GCTGAGATCTACAGACCATCCAGTCTTGGCCAGGCGCTGAGAGACCTGATGCT 945
940 GCTGAGATCTACAGACCATCCAGTCTTGGCCAGGCGCTGAGAGACCTGATGCT 999
946 TGTGCGCTATGAGAGACCTGGCTGAGCGCCGTGGCCAGACTCCCGAATGTATGAT 1005
1000 TGTGCGCTATGAGAGACCTGGCTGAGCGCCGTGGCCAGACTCCCGAATGTATGAT 1059
1006 CGTGGATTTGGAATTTCTGCCCATCTTTCAGACCTGGGTCATTAATCAACCCGAGCAA 1065
1060 CGTGGATTTGGAATTTCTGCCCATCTTTCAGACCTGGGTCATTAATCAACCCGAGCAA 1119
1066 GGGCATGGTGCACACGCTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGC 1125
1120 GGGCATGGTGCACACGCTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGC 1179
1126 TTTGGCGCTGGCTTTGGCCATATGAAAGGTTTCTGATTCAGAAACCTGGTGGCATGC 1185
1180 TTTGGCGCTGGCTTTGGCCATATGAAAGGTTTCTGATTCAGAAACCTGGTGGCATGC 1239
1186 CATGAATTTCTGCTGGCTACCGCCAGCTCAGATCTGAACAAGAGAAACCTGTTGCT 1245
1240 CATGAATTTCTGCTGGCTACCGCCAGCTCAGATCTGAACAAGAGAAACCTGTTGCT 1299
1246 GGAATCTTCTGCTACCTGACAGTCTCCCTGAGCAATCCACTAGAGGTTGAGAGGCTT 1305
1300 GGAATCTTCTGCTACCTGACAGTCTCCCTGAGCAATCCACTAGAGGTTGAGAGGCTT 1359
1306 TGTGCGACCTGGTGTAGAGCTGACGCTGCTCTGATGCTTCTGAGCGTTCCTACA 1365
1360 TGTGCGACCTGGTGTAGAGCTGACGCTGCTCTGATGCTTCTGAGCGTTCCTACA 1419
1366 TCTCTGAGCCTTAATCAATGCTGTGGTATCAACACTGATGATGATGATGATGATGATG 1425
1420 TCTCTGAGCCTTAATCAATGCTGTGGTATCAACACTGATGATGATGATGATGATGATG 1479
1426 TGTCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGAACT 1485
1480 TGTCTCAAGCAGAGAGACTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGAACT 1539
1486 TATGTGAGCAGACATCCACACAGTGAAGAGGATGATGCTCTCTCTCTCTCTCTCTCT 1545
1540 TATGTGAGCAGACATCCACACAGTGAAGAGGATGATGCTCTCTCTCTCTCTCTCTCT 1599
1546 TCTCTGTGGGAGACTTTCAGAGACTTTGTGGCCTGAGGCGCTATTAAGCAGCAGAGT 1605
1600 TCTCTGTGGGAGACTTTCAGAGACTTTGTGGCCTGAGGCGCTATTAAGCAGCAGAGT 1659
1606 ATCAGTGAATTTGATCAATAAACCCTCCTGTCACATCTTGCCCAATGGGGAATGATCT 1665
1660 ATCAGTGAATTTGATCAATAAACCCTCCTGTCACATCTTGCCCAATGGGGAATGATCT 1719
1666 TTTCACCAAGAGCTCACCAGATTTTCCACAGAGATGC 1703
1720 TTTCACCAAGAGCTCACCAGATTTTCCACAGAGATGC 1757

```

```

RESULT 3
US-09-786-240-26
: Sequence 26: Application US/09786240
: Patent No. 6558935
: GENERAL INFORMATION:
: APPLICANT: INCYTE PHARMACEUTICALS, INC.
: APPLICANT: TANG, Y. Tom
: APPLICANT: CORLEY, Neil C.
: APPLICANT: GUEGLER, Karl J.
: APPLICANT: BAUGHN, Mariah R.
: APPLICANT: LAL, Preeti
: APPLICANT: YUE, Henry
: APPLICANT: HILMAN, Jennifer L.
: APPLICANT: AZIMZAI, Yalda
: TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
: FILE REFERENCE: PF-0592 PCT
: CURRENT APPLICATION NUMBER: US/09/786, 240
: PRIOR FILING DATE: 2002-03-12
: PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,6
: PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11
: NUMBER OF SEQ ID NOS: 33
: SOFTWARE: PERL Program
: SEQ ID NO 26
: LENGTH: 2065
: TYPE: DNA
: ORGANISM: Homo sapiens
: FEATURE:
: NAME/KEY: misc-feature
: OTHER INFORMATION: Incyte ID No. 6558935 2617407CB1
US-09-786-240-26

```

Query Match 56.4%; Score 1153; DB 4; Length 2065;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

```

QY 106 CAAGTCTTCCACTTTCAGACACAAATGCTACTGCTTAAATAAATGAAGCTCTGCTTCT 165
DB 152 CAAGTCTTCCACTTTCAGACACAAATGCTACTGCTTAAATAAATGAAGCTCTGCTTCT 211
QY 166 GGTTCCTCCAGATGGCCATCTTGGCTATTTCTTCAATGATGACGCCACAAATCACTG 225
DB 212 GGTTCCTCCAGATGGCCATCTTGGCTATTTCTTCAATGATGACGCCACAAATCACTG 271
QY 226 CCGTCTATGAAGCAGACGCCGAGCGCATGACGCTGCTGCTTCTCTCTGCGGCTC 285
DB 272 CCGTCTATGAAGCAGACGCCGAGCGCATGACGCTGCTGCTTCTCTCTGCGGCTC 331
QY 286 TGGCTCTTCTTGTGGGAGCTTTTGGGACACACCAAGATTTTCTACCTGATGA 345
DB 332 TGGCTCTTCTTGTGGGAGCTTTTGGGACACACCAAGATTTTCTACCTGATGA 391
QY 346 GCCCGCTGGACGCTGTGATGACCTTCAAGCAGACCGCTGATGCTGACATGCG 405
DB 392 GCCCGCTGGACGCTGTGATGACCTTCAAGCAGACCGCTGATGCTGACATGCG 451
QY 406 TGTGCGGATGCTATGAGGCGGCTTGTGTGACATGAGGCTTGTGATGCTACAT 465
DB 452 TGTGCGGATGCTATGAGGCGGCTTGTGTGACATGAGGCTTGTGATGCTACAT 511
QY 466 GGAACCTGTGCTCCCGGAGACAGTCCAGCTCTTTCAGTGGAGAAACAGCGGGCGCTGTG 525
DB 512 GGAACCTGTGCTCCCGGAGAGAGTCCAGCTCTTTCAGTGGAGAAACAGCGGGCGCTGTG 571
QY 526 TTTCTGACCTGCTGCTGATCATCCCAAGATGAATATATCCCCGGGCTCACTGAG 585
DB 572 TTTCTGACCTGCTGCTGATCATCCCAAGATGATATATCCCCGGGCTCACTGAG 630
QY 586 GCTCCTGTGAGTCAAGAGCGCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 644
DB 631 GCTCCTGTGAGTCAAGAGCGCTTGTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 690
QY 645 ACCTGTGCTCAAGAGAGTGGCTTTTCAACCTGACGCTCTTACCGGCTGCTGAAG 704

```

Db 691 AGGTGCTGCTCAAGAGAGTGCCTGCTTCTCAACCTGACGTCCCTCTACCCGCTGCTMAAG 750
QY ACCCTCCCTCAACCTGATATGTCACCTGCTGCGGAGACCCCGGCGCTGCTCCGTT 764
Db 751 ACCCTCCCTCAACCTGATATGTCACCTGCTGCGGAGACCCCGGCGCTGCTCCGTT 810
QY 765 CCCGAGAGAGCAAGAGAGAGATCTGATGATGACAGTGCATTTGATGAGGAGCAGATG 824
Db 811 CCCGAGAGAGCAAGAGAGAGATCTGATGATGACAGTGCATTTGATGAGGAGCAGATG 870
QY 825 ACCAAAACTCAAGAGAGAGAGACACCTACTATGATGACAGTGCATTTGACCAAGCC 884
Db 871 ACCAGAACTCAAGAGAGAGAGACACCTACTATGATGACAGTGCATTTGACCAAGCC 930
QY 885 AGCTGAGATCTCAAGAGAGAGATCTGATGACAGTGCATTTGACCAAGCC 944
Db 931 AGCTGAGATCTCAAGAGAGAGATCTGATGACAGTGCATTTGACCAAGCC 990
QY 945 TTGTGCGCTATGAGAGAGCTGCTGAGCCCTGTGCGGAGAGAGAGAGATGATGAT 1004
Db 991 TTGTGCGCTATGAGAGAGCTGCTGAGCCCTGTGCGGAGAGAGAGATGATGAT 1050
QY 1005 TGTGAGATGAGATCTGCTGAGCCCTGATGACAGTGCATTTGACCAAGCC 1064
Db 1051 TGTGAGATGAGATCTGCTGAGCCCTGATGACAGTGCATTTGACCAAGCC 1110
QY 1065 AGGCGATGAGAGAGAGAGAGAGATCTGATGACAGTGCATTTGACCAAGCC 1124
Db 1111 AGGCGATGAGAGAGAGAGAGAGATCTGATGACAGTGCATTTGACCAAGCC 1170
QY 1125 CTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1184
Db 1171 CTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1230
QY 1185 CCATGATTTGCTGAGGCTGCTGAGCCGATGATGATGATGATGATGATGATGATGATG 1244
Db 1231 CCATGATTTGCTGAGGCTGCTGAGCCGATGATGATGATGATGATGATGATGATGATG 1290
QY 1245 TGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1304
Db 1291 TGGATTTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1350
QY 1305 TTGCTGCGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1364
Db 1351 TTGCTGCGCAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1410
QY 1365 ATCTGAGACCTTAATACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1424
Db 1411 ATCTGAGACCTTAATACATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1470
QY 1425 GTGCTCAAGCAGAGAGAGCTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1484
Db 1471 GTGCTCAAGCAGAGAGAGCTTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1530
QY 1485 TTAATGAGCAGACATCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1544
Db 1531 TTAATGAGCAGACATCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1590
QY 1545 TTCTGCTGAGCAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1604
Db 1591 TTCTGCTGAGCAGAGCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1650
QY 1605 TATCAGTGAATGATCATTAACCTCCCTGCTCAGATCTTGCCTAATGGGAGATGATC 1664
Db 1651 TATCAGTGAATGATCATTAACCTCCCTGCTCAGATCTTGCCTAATGGGAGATGATC 1710
QY 1665 TTTCAACCAAG 1703
Db 1711 TTTCAACCAAG 1749

RESULT 4
US-08-992-334-1

Sequence 1, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
TITLE OF INVENTION: PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
PRIORITY APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: FR 932/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ. ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHEtical: NO
ANTI-SENSE: YES
IMMEDIATE SOURCE:
CLONE: pc+host4
US-08-992-334-1
Query Match 0.9%; Score 19; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Db 1821 TCTCTGACAAAGAGCAAG 1839
1818 TCTCTGACAAAGAGCAAG 1836
RESULT 5
US-08-302-752-1
Sequence 1, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSITIVE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-1

Query Match 0.9%: Score 19; DB 3; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCCTGACAAAGAGCAAG 1839
 ||||||||||||||||
DB 1818 TCCTGACAAAGAGCAAG 1836

RESULT 6
US-08-992-334-2
Sequence 2, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
TITLE OF INVENTION: PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR FR92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93.31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-2

Query Match 0.9%: Score 19; DB 2; Length 5234;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCCTGACAAAGAGCAAG 1839
 ||||||||||||||||
DB 3260 TCCTGACAAAGAGCAAG 3278

RESULT 7
US-08-302-752-2
Sequence 2, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-2

Query Match 0.9%: Score 19; DB 3; Length 5234;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCCTGACAAAGAGCAAG 1839
 ||||||||||||||||
DB 3260 TCCTGACAAAGAGCAAG 3278

RESULT 8
US-08-992-334-3
Sequence 3, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
TITLE OF INVENTION: PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California

COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 932/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-8900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6722 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-3

Query Match
Best Local Similarity 100.0%; Score 19; DB 2; Length 6722;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCTCTGACAAAGAGCAG 1839
|||||
DB 4748 TCTCTGACAAAGAGCAG 4766

RESULT 9
S-08-302-752-3
Sequence 3, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6722 base pairs
TYPE: nucleic acid
STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-3

Query Match
Best Local Similarity 100.0%; Score 19; DB 3; Length 6722;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCTCTGACAAAGAGCAG 1839
|||||
DB 4748 TCTCTGACAAAGAGCAG 4766

RESULT 10
US-09-020-956-69
Sequence 69, Application US/09020956
Patent No. 6261562
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS
NUMBER OF SEQUENCES: 178
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/020,956
FILING DATE: 09-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Makl, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 536 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: CDNA
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-09-020-956-69

Query Match
Best Local Similarity 100.0%; Score 18; DB 3; Length 536;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCATTGCTTGGG 18
|||||
DB 18 GAATTCATTGCTTGGG 35

RESULT 11
US-09-030-607-69
Sequence 69, Application US/09030607
Patent No. 6262245
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, David C.
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE CANCER AND METHODS

NUMBER OF SEQUENCES: 224
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED AND BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/030,607
FILING DATE: 25-FEB-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: MAKI, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.427C3
TELEPHONE: (206) 622-4900
TELEFAX: (206) 682-6031
INFORMATION FOR SEQ ID NO: 69:
SEQUENCE CHARACTERISTICS:
LENGTH: 536 base pairs
TYPE: nucleic acid
STRADEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cdna
ORIGINAL SOURCE:
ORGANISM: Homo sapiens
US-09-030-607-69

Query Match 0.9%; Score 18; DB 3; Length 536;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAATTCATGTGTGGG 18
DB 18 GAATTCATGTGTGGG 35

RESULT 12
US-09-439-313-69
Sequence 69, Application US/09439313
Patent No. 6329505
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang Yuqul
APPLICANT: Reed, Steven G.
APPLICANT: Kalos, Michael
APPLICANT: Fanger, Gary
APPLICANT: Retter, Mark
APPLICANT: Solk, John
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
FILE REFERENCE: 210121.427C9
CURRENT APPLICATION NUMBER: US/09/439,313
NUMBER OF SEQ ID NOS: 575
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 69
LENGTH: 536
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc-feature

LOCATION: (1)...(536)
OTHER INFORMATION: n - A,T,C or G
US-09-439-313-69

Query Match 0.9%; Score 18; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAATTCATGTGTGGG 18
DB 18 GAATTCATGTGTGGG 35

RESULT 13
US-09-352-616A-69
Sequence 69, Application US/09352616A
Patent No. 6395278
GENERAL INFORMATION:
APPLICANT: Dillon, Davin C.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang, Yuqul
APPLICANT: Xu, Jiangchun
APPLICANT: Mitcham, Jennifer Lynn
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY AND DIAGNOSIS
FILE REFERENCE: 210121.427C8
CURRENT APPLICATION NUMBER: US/09/352,616A
CURRENT FILING DATE: 1999-07-13
NUMBER OF SEQ ID NOS: 472
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 69
LENGTH: 536
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)...(536)
OTHER INFORMATION: n - A,T,C or G
US-09-352-616A-69

Query Match 0.9%; Score 18; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 GAATTCATGTGTGGG 18
DB 18 GAATTCATGTGTGGG 35

RESULT 14
US-09-232-149A-69
Sequence 69, Application US/09232149A
Patent No. 6465611
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer Lynn
TITLE OF INVENTION: COMPOUNDS FOR IMMUNOTHERAPY OF PROSTATE
FILE REFERENCE: 210121.427C6
CURRENT APPLICATION NUMBER: US/09/232,149A
CURRENT FILING DATE: 1999-01-15
NUMBER OF SEQ ID NOS: 338
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 69
LENGTH: 536
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc-feature
LOCATION: (1)...(536)
OTHER INFORMATION: n - A,T,C or G
US-09-232-149A-69

Job time : 145 secs

Query Match 0.9%; Score 18; DB 4; Length 536;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 GAATTCATTGTTGGG 18
DB 18 GAATTCATTGTTGGG 35

RESULT 15

US-09-107-532A-1333

; Sequence 1333 Application US/09107532A

; Patent No. 6583275

; GENERAL INFORMATION:

; APPLICANT: Lynn A Doucette-Stamm and David Bush

; TITLE OF INVENTION: NUCLEIC ACID SEQUENCES RELATING TO
ENTEROCOCCUS FAECIUM FOR DIAGNOSTICS AND THERAPEUTICS

; NUMBER OF SEQUENCES: 7310

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: GENOME THERAPEUTICS CORPORATION

; STREET: 100 Beaver Street

; CITY: Waltham

; STATE: Massachusetts

; COUNTRY: USA

; ZIP: 02354

; COMPUTER READABLE FORM:

; MEDIUM TYPE: CD-ROM ISO9660

; COMPUTER: PC

; OPERATING SYSTEM: <Unknown>

; SOFTWARE: ASCII

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/09/107/532A

; FILING DATE: 30-Jun-1998

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 60/085,598

; FILING DATE: 14 May 1998

; APPLICATION NUMBER: 60/051571

; FILING DATE: July 2, 1997

; ATTORNEY/AGENT INFORMATION:

; NAME: Ariunello, Pamela Deneke

; REGISTRATION NUMBER: 40,489

; REFERENCE/DOCKET NUMBER: CTC-012

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: (781)893-5007

; TELEFAX: (781)893-8277

; INFORMATION FOR SEQ ID NO: 1333:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 699 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: double

; TOPOLOGY: circular

; MOLECULE TYPE: DNA (genomic)

; HYPOTHETICAL: NO

; ANTI-SENSE: NO

; ORIGINAL SOURCE:

; ORGANISM: Enterococcus faecium

; FEATURE:

; NAME/KEY: misc_feature

; LOCATION: (B) LOCATION 1...699

; SEQUENCE DESCRIPTION: SEQ ID NO: 1333:

US-09-107-532A-1333

Query Match 0.9%; Score 18; DB 4; Length 699;
Best Local Similarity 100.0%; Pred. No. 41;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1528 TTCTCTTTCTTGATCT 1545
DB 473 TTCTCTTTCTTGATCT 490

Search completed: August 15, 2003, 08:09:08


```

Db      |||||
340 TGGCTCTCTTTTGGGGGACCTTTTGGGAGACCCAGATGTTTCTTACTGTATGA 399
QY      GCGCGCGTGGACAGTGTGATGACCTTCAAGACAGACACCGCTGGATCTGTGACATGGC 405
Db      |||||
400 GCGCGCGTGGACAGTGTGATGACCTTCAAGACAGACACCGCTGGATCTGTGACATGGC 459
QY      TGTGGGGGATCTGATACGCGGCGTCTTGTGTGCGACATGAGCGTCTTGTATGCCATAT 465
Db      |||||
460 TGTGGGGGATCTGATACGCGGCGTCTTGTGTGCGACATGAGCGTCTTGTATGCCATAT 519
QY      GGAACCTGGTCCCGGAGACATCCAGCCTTTTCAAGTGGGAGAACAGCGCGCGCTGTG 525
Db      |||||
520 GGAACCTGGTCCCGGAGACATCCAGCCTTTTCAAGTGGGAGAACAGCGCGCGCTGTG 579
QY      TTCTGACACCTGCTGTGACATCATCCACAAATGATGAATCATCCCGGGCTCACTGACG 585
Db      |||||
580 TTCTGACACCTGCTGTGACATCATCCACAAATGATGAATCATCCCGGGCTCACTGACG 639
QY      GGTCTCTGTGACATCAACAGCCCTTTGAGTGTGGAGAGAGCGCTGCCCTCTACAGCA 645
Db      |||||
640 GGTCTCTGTGACATCAACAGCCCTTTGAGTGTGGAGAGAGCGCTGCCCTCTACAGCA 699
QY      CGTGGGCTCAAGAGAGTGGCTCTTCAACCTGAGTCCCTACCGGCTGTGAGAGA 705
Db      |||||
700 CGTGGGCTCAAGAGAGTGGCTCTTCAACCTGAGTCCCTACCGGCTGTGAGAGA 759
QY      CCGCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGCGGCGCTGTGCGTTC 765
Db      |||||
760 CCGCTCCCTCAACCTGATATGTCACCTGTCGGGAGCCCGCGGCGCTGTGCGTTC 819
QY      CCGAGAGCGCAAGAGAGATCTCATGTTGACAGTGTGATGATGATGATGATGATGATGAT 825
Db      |||||
820 CCGAGAGCGCAAGAGAGATCTCATGTTGACAGTGTGATGATGATGATGATGATGATGAT 879
QY      GCAAAAACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 885
Db      |||||
880 GCAAAAACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 939
QY      GGTGAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 945
Db      |||||
940 GGTGAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 999
QY      TGTGGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1005
Db      |||||
1000 TGTGGGCTTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1059
QY      CGTGGGATTTGAATCTTGGCCCATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1065
Db      |||||
1060 CGTGGGATTTGAATCTTGGCCCATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1119
QY      GGGGATTTGGTGAACAGAGCTTTCCACACAATGCCAGAGATGCCCTTATGTCTCCAGGC 1125
Db      |||||
1120 GGGGATTTGGTGAACAGAGCTTTCCACACAATGCCAGAGATGCCCTTATGTCTCCAGGC 1179
QY      TTGGGCGTGTGCTTGGCCCTATGAAAAGGTTCTGAGCTTCAAAAAGCCTGTGGGATGC 1185
Db      |||||
1180 TTGGGCGTGTGCTTGGCCCTATGAAAAGGTTCTGAGCTTCAAAAAGCCTGTGGGATGC 1239
QY      CATGAATTTGCTGGGCTACCGGACAGTGTGAACAAAGAGAGAGAGAGAGAGAGAGAG 1245
Db      |||||
1240 CATGAATTTGCTGGGCTACCGGACAGTGTGAACAAAGAGAGAGAGAGAGAGAGAGAG 1299
QY      GGAATCTTGTCTACCTGAGCTGTCCCTGAGCAATCCACTAAGAGAGAGAGAGAGAG 1305
Db      |||||
1300 GGAATCTTGTCTACCTGAGCTGTCCCTGAGCAATCCACTAAGAGAGAGAGAGAGAG 1359
QY      TGTGCGACAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTG 1365
Db      |||||
1360 TGTGCGACAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTGAGCTGTG 1419
QY      TGTGCGACCTTAACTACATGTCTGTGGTATCACACTGATGTGATGTGTGTCCACAG 1425

```

```

Db      1420 TGTGAGAGCTTAACTACATGTCTGTGGTATCACACTGATGTGATGTGTCCACAG 1479
QY      1426 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1485
Db      1480 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1539
QY      1486 TATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1545
Db      1540 TATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1599
QY      1546 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1605
Db      1600 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659
QY      1606 ATCAGTGAATTTGATCCATTAACCTCCCTGTCCATATCTTGGCCATGGGAGATGAT 1665
Db      1660 ATCAGTGAATTTGATCCATTAACCTCCCTGTCCATATCTTGGCCATGGGAGATGAT 1719
QY      1666 TTACCAAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
Db      1720 TTACCAAAAGAGCTCACAGCATTTTCCACAGAGATGC 1757

RESULT 2
US-10-007-262-2
; Sequence 2, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007,262
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-007-262-2

Query Match      75.7%; Score 1547; DB 13; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY      106 CAAAGTCTTCCACTTCAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTTTCT 165
Db      160 CAAAGTCTTCCACTTCAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTTTCT 219
QY      166 GGTTCGCCAGATGGGCGATTTGGCTCTATCTTCCACATGTACAGGACAAATCAGCTC 225
Db      220 GGTTCGCCAGATGGGCGATTTGGCTCTATCTTCCACATGTACAGGACAAATCAGCTC 279
QY      226 CCTGTCTATGAAGGACACAGCCCGAGGCGATGACAGTGTGCTGTCTTCTGCGGCTC 285
Db      280 CCTGTCTATGAAGGACACAGCCCGAGGCGATGACAGTGTGCTGTCTTCTGCGGCTC 339
QY      286 TGGCTCTCTTTTGTGGGGACAGCTTTTGGGAGACAGCCAGATGTTTCTTACTGTATGA 345
Db      340 TGGCTCTCTTTTGTGGGGACAGCTTTTGGGAGACAGCCAGATGTTTCTTACTGTATGA 399
QY      346 GCCCGCTGGACAGTGTGATGACCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405
Db      400 GCCCGCTGGACAGTGTGATGACCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 459
QY      406 TGTGCGGATCTGATACGCGGCGTCTTGTGTGCGACATGAGCGTCTTGTATGCCATAT 465
Db      460 TGTGCGGATCTGATACGCGGCGTCTTGTGTGCGACATGAGCGTCTTGTATGCCATAT 519

```


RESULT 8
US-09-960-352-11099/C
; Sequence 11099, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengping
; APPLICANT: Batt, John C.
; APPLICANT: Mathalaagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; TITLE OF INVENTION: MUSCLE AND FAT DEPOSITION
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/360,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11099
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-BOVMS1-008-Q1-E1-DB
US-09-960-352-11099

Query Match 1.0%; Score 20; DB 10; Length 398;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 175 GATGCCATCTTGGCTCAT 194
DB 240 GATGCCATCTTGGCTCAT 221

RESULT 9
US-10-027-632-140660
; Sequence 140660, Application US/10027632
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; TITLE OF INVENTION: Polymorphisms in the Human Genome
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140660
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-140660

Query Match 1.0%; Score 20; DB 13; Length 570;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 700 GAAGACCCCTCCCTCAACC 719
DB 273 GAAGACCCCTCCCTCAACC 292

RESULT 10
US-10-205-823-334/C
; Sequence 334, Application US/10205823
; Publication No. US20030108963A1
; GENERAL INFORMATION:
; APPLICANT: Schlegel, Robert
; APPLICANT: Monahan, John E.
; APPLICANT: Endege, Wilson O.
; APPLICANT: Gannavarapu, Manjula
; APPLICANT: Gorbacheva, Bella
; APPLICANT: Hoersch, Sebastian
; APPLICANT: Kamatkar, Shubhangi
; APPLICANT: Monsey, Angela M.
; APPLICANT: Glatz, Karen
; APPLICANT: Zhao, Xumei
; APPLICANT: Anderson, Dustin
; TITLE OF INVENTION: NOVEL GENES, COMPOSITIONS, KITS, AND
; TITLE OF INVENTION: METHODS FOR IDENTIFICATION, ASSESSMENT, PREVENTION, AND
; TITLE OF INVENTION: THERAPY OF PROSTATE CANCER
; FILE REFERENCE: MRI-044
; CURRENT APPLICATION NUMBER: US/10/205,823
; CURRENT FILING DATE: 2002-07-25
; PRIOR APPLICATION NUMBER: 60/307,982
; PRIOR FILING DATE: 2001-07-25
; PRIOR APPLICATION NUMBER: 60/314,356
; PRIOR FILING DATE: 2001-08-22
; PRIOR APPLICATION NUMBER: 60/325,020
; PRIOR FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: 60/341,746
; PRIOR FILING DATE: 2001-12-12
; PRIOR APPLICATION NUMBER: 60/362,158
; PRIOR FILING DATE: 2002-03-05
; NUMBER OF SEQ ID NOS: 455
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 334
; LENGTH: 11102
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-205-823-334

Query Match 1.0%; Score 20; DB 14; Length 11102;
Best Local Similarity 100.0%; Pred. No. 8.9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1523 TGCTCTCTCTCTTCTTGA 1542
DB 722 TGCTCTCTCTCTTCTTGA 703

RESULT 11
US-10-003-806-6
; Sequence 6, Application US/10003806
; Publication No. US20020119929A1
; GENERAL INFORMATION:
; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulunk, Alexander I.
; APPLICANT: Zhu, Qichao
; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P020660S1/10024824
; CURRENT APPLICATION NUMBER: US/10/003,806
; CURRENT FILING DATE: 2001-11-02
; PRIOR APPLICATION NUMBER: 60/245,872
; PRIOR FILING DATE: 2000-11-03
; NUMBER OF SEQ ID NOS: 14
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 6
; LENGTH: 180557
; TYPE: DNA
; ORGANISM: Human
US-10-003-806-6

Query Match 1.0%; Score 20; DB 13; Length 180557;

Best Local Similarity 100.0%; Pred. No. 8.7;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1893 TTTTCCATTACATAGAAA 1912
|||||
DB 2567 TTTTCCATTACATAGAAA 2586

RESULT 12
US-10-003-806-9

; Sequence 9, Application US/10003806
; Publication No. US20020119929A1
; GENERAL INFORMATION:

; APPLICANT: Bishop, Colin E.
; APPLICANT: Agoulouk, Alexander I.

; APPLICANT: Zhu, Qichao

; TITLE OF INVENTION: CAN 1 AND ITS ROLE IN MAMMALIAN INFERTILITY
; FILE REFERENCE: P02066US1/10024824

; CURRENT APPLICATION NUMBER: US/10/003,806

; CURRENT FILING DATE: 2001-11-02

; PRIOR APPLICATION NUMBER: 60/245,872

; PRIOR FILING DATE: 2000-11-03

; NUMBER OF SEQ ID NOS: 14

; SOFTWARE: PatentIn version 3.1

; SEQ ID NO 9

; LENGTH: 180557

; TYPE: DNA

; ORGANISM: Human

US-10-003-806-9

Query Match
Best Local Similarity 100.0%; Score 20; DB 13; Length 180557;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1893 TTTTCCATTACATAGAAA 1912
|||||
DB 2567 TTTTCCATTACATAGAAA 2586

RESULT 13

US-09-867-701-8841/C

; Sequence 8841, Application US/09867701

; Patent No. US20020132237A1

; GENERAL INFORMATION:

; APPLICANT: Aglate, Paul A.

; APPLICANT: Jones, Robert

; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.497

; CURRENT APPLICATION NUMBER: US/09/867,701

; CURRENT FILING DATE: 2001-05-29

; NUMBER OF SEQ ID NOS: 10912

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8841

; LENGTH: 341

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-867-701-8841

Query Match
Best Local Similarity 100.0%; Score 19; DB 10; Length 341;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACAGACGAAGGTAG 41
|||||
DB 330 GAAGACAGACGAAGGTAG 312

RESULT 14

US-09-867-701-8888/C

; Sequence 8888, Application US/09867701

; Patent No. US20020132237A1

; GENERAL INFORMATION:

; APPLICANT: Aglate, Paul A.

; APPLICANT: Jones, Robert

; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY

; FILE REFERENCE: 210121.497

; CURRENT APPLICATION NUMBER: US/09/867,701

; CURRENT FILING DATE: 2001-05-29

; NUMBER OF SEQ ID NOS: 10912

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 8888

; LENGTH: 341

; TYPE: DNA

; ORGANISM: Homo sapien

US-09-867-701-8888

Query Match
Best Local Similarity 100.0%; Score 19; DB 10; Length 341;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACAGACGAAGGTAG 41
|||||
DB 330 GAAGACAGACGAAGGTAG 312

RESULT 15

US-09-759-143-649

; Sequence 649, Application US/09759143

; Patent No. US2002022248A1

; GENERAL INFORMATION:

; APPLICANT: Xu, Jianshun

; APPLICANT: Dillon, Davin C.

; APPLICANT: Mitcham, Jennifer L.

; APPLICANT: Harlocker, Susan L.

; APPLICANT: Jiang, Yugu

; APPLICANT: Henderson, Robert A.

; APPLICANT: Kalos, Michael D.

; APPLICANT: Fanger, Gary R.

; APPLICANT: Retter, Marc W.

; APPLICANT: Stolk, John A.

; APPLICANT: Day, Craig H.

; APPLICANT: Vedvick, Thomas S.

; APPLICANT: Carter, Darrick

; APPLICANT: Li, Samuel

; APPLICANT: Wang, Aijun

; APPLICANT: Skelky, Yasir A.W.

; APPLICANT: Hepler, William

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND

; FILE REFERENCE: 210121.427C23

; CURRENT APPLICATION NUMBER: US/09/759,143

; CURRENT FILING DATE: 2001-01-12

; NUMBER OF SEQ ID NOS: 934

; SOFTWARE: FastSeq for Windows Version 3.0

; SEQ ID NO 649

; LENGTH: 349

; TYPE: DNA

; ORGANISM: Homo sapien

; FEATURE:

; NAME/KEY: misc-feature

; LOCATION: (1)...(349)

; OTHER INFORMATION: n = A,T,C or G

US-09-759-143-649

Query Match
Best Local Similarity 100.0%; Score 19; DB 9; Length 349;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 GAATTCATTGTGTGGGT 19
|||||
DB 72 GAATTCATTGTGTGGGT 90

Fri Aug 15 08:09:28 2003

us-09-645-078-1.rnpb

Page 7

Search completed: August 15, 2003, 08:06:32
Job time : 539 secs

THIS PAGE BLANK (USPTO)

[illegible]

XX Human and murine glycosyl sulfotransferase 3 and related
 PT polynucleotides
 XX
 PS Claim 2; Fig 1; 59pp; English.

CC This sequence is the human glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.

XX Sequence 386 AA;

Query Match 100.0%; Score 2038; DB 20; Length 386;
 Best Local Similarity 100.0%; Pred. No. 1.3e-205;
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLTPKMKLLFLVSQMAITLFFHMYSHNSSLKMAQPERMHVLTSSMRGSSFPVQ 60
 DB 1 MLTPKMKLLFLVSQMAITLFFHMYSHNSSLKMAQPERMHVLTSSMRGSSFPVQ 60
 QY 61 LFGQHPDVFYLMRPANHWMTFKQSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 DB 61 LFGQHPDVFYLMRPANHWMTFKQSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 QY 121 SSLFQWNSRALCSAPACDIIPODEIIPRAHCHLCSQPFEEVEKACRSYSHVLEVR 180
 DB 121 SSLFQWNSRALCSAPACDIIPODEIIPRAHCHLCSQPFEEVEKACRSYSHVLEVR 180
 QY 181 FFMLOSLYPLKDPSPSLNHLVLRDPRAVRSRERTKGLMDSRIYMGQHOKLKED 240
 DB 181 FFMLOSLYPLKDPSPSLNHLVLRDPRAVRSRERTKGLMDSRIYMGQHOKLKED 240
 QY 241 QPYVMQVTCOSOLEIYKTQSLPKALOERYLLVRYEDLARAVAOVSRYEFGLEFLP 300
 DB 241 QPYVMQVTCOSOLEIYKTQSLPKALOERYLLVRYEDLARAVAOVSRYEFGLEFLP 300
 QY 301 HLOTWVHNITRGKMGDHAFTNARDALNVSAWRMSLPYEKYSRLQKACGDMNLGTR 360
 DB 301 HLOTWVHNITRGKMGDHAFTNARDALNVSAWRMSLPYEKYSRLQKACGDMNLGTR 360
 QY 361 HVRSQEQQRNLLDLSTWTPPEQIH 386
 DB 361 HVRSQEQQRNLLDLSTWTPPEQIH 386

RESULT 2
 AAM93309 standard; Protein; 386 AA.

XX AAM93309;
 XX
 AC 06-NOV-2001 (first entry)
 XX
 DT Human polypeptide, SEQ ID NO: 2817.
 XX
 DE Human; full length cDNA; cDNA synthesis; oligo-capping.
 KW
 XX Homo sapiens.
 OS
 XX EPI130094-A2.
 PN
 XX 05-SEP-2001.
 PD
 XX 07-JUL-2000; 2000EP-0114089.
 PF
 XX 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

PA Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y.
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX WPI: 2001-524255/58.
 DR N-PSDB; AAK94229.

PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -

PS Claim 8; SEQ ID NO 2817; 1380pp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a polypeptide
 CC encoded by a full length human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 386 AA;

Query Match 99.9%; Score 2035; DB 22; Length 386;
 Best Local Similarity 99.7%; Pred. No. 2.7e-205;
 Matches 385; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLTPKMKLLFLVSQMAITLFFHMYSHNSSLKMAQPERMHVLTSSMRGSSFPVQ 60
 DB 1 MLTPKMKLLFLVSQMAITLFFHMYSHNSSLKMAQPERMHVLTSSMRGSSFPVQ 60
 QY 61 LFGQHPDVFYLMRPANHWMTFKQSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 DB 61 LFGQHPDVFYLMRPANHWMTFKQSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 QY 121 SSLFQWNSRALCSAPACDIIPODEIIPRAHCHLCSQPFEEVEKACRSYSHVLEVR 180
 DB 121 SSLFQWNSRALCSAPACDIIPODEIIPRAHCHLCSQPFEEVEKACRSYSHVLEVR 180
 QY 181 FFMLOSLYPLKDPSPSLNHLVLRDPRAVRSRERTKGLMDSRIYMGQHOKLKED 240
 DB 181 FFMLOSLYPLKDPSPSLNHLVLRDPRAVRSRERTKGLMDSRIYMGQHOKLKED 240
 QY 241 QPYVMQVTCOSOLEIYKTQSLPKALOERYLLVRYEDLARAVAOVSRYEFGLEFLP 300
 DB 241 QPYVMQVTCOSOLEIYKTQSLPKALOERYLLVRYEDLARAVAOVSRYEFGLEFLP 300
 QY 301 HLOTWVHNITRGKMGDHAFTNARDALNVSAWRMSLPYEKYSRLQKACGDMNLGTR 360
 DB 301 HLOTWVHNITRGKMGDHAFTNARDALNVSAWRMSLPYEKYSRLQKACGDMNLGTR 360
 QY 361 HVRSQEQQRNLLDLSTWTPPEQIH 386
 DB 361 HVRSQEQQRNLLDLSTWTPPEQIH 386

RESULT 3
 AAU11274 standard; Protein; 380 AA.

XX AAU11274;
 XX
 AC 12-MAR-2002 (first entry)
 XX
 DT Human L-selectin sulfotransferase-2 (LSS-2) protein.
 XX
 DE Human; beta1,3-galactosyltransferase; MECA-79;
 KW

KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
 KW allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
 KW delayed-type hypersensitivity reaction; hyperplastic thymus; antilucer;
 KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;
 KW antiallergic.
 XX Homo sapiens.
 XX OS
 XX MO200185177-A1.
 XX PN
 XX 15-NOV-2001.
 XX PD
 XX 10-MAY-2001; 2001WO-US15452.
 XX PE
 XX 11-MAY-2000; 2000US-0569320.
 XX PR
 XX (BURN-) BURNHAM INST.
 XX PA
 XX Fukuda M, Yeh J, Hiraoka N;
 XX PI
 XX WPI: 2002-075226/10.
 XX DR
 XX N-PSDB; AAS16947.
 XX PS
 XX New enzyme, useful for modifying acceptor molecule, comprises an
 PT isolated L-selectin sulfotransferase-2 that directs expression of
 PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
 PT intestinal G1CNC 6-sulfotransferase
 XX
 XX Claim 21; Fig 4; 98pp; English.
 XX
 CC The present invention provides a method of modifying an acceptor molecule
 CC by contacting the acceptor with an isolated
 CC betal,3-N-acetylglucosaminyltransferase (betal,3gnt) or an active
 CC fragment, where betal,3gnt directs expression of a MECA-79 antigen. The
 CC invention also provides a method of treating or preventing an
 CC L-selectin-mediated condition by reducing the expression or activity of a
 CC betal,3gnt that directs expression of a MECA-79 antigen. This can be done
 CC by administering to the subject an oligosaccharide L-selectin antagonist
 CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
 CC administering antibody material that specifically binds betal,3gnt,
 CC and/or a betal,3gnt antisense nucleic acid molecule. L-selectin
 CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
 CC Alternatively, the expression or activity of LST-2 or its active
 CC fragment can be reduced in combination with reducing the expression or
 CC activity of betal,3gnt. The method is useful for treating L-selectin
 CC mediated conditions such as Crohn's disease and ulcerative colitis,
 CC inflammatory disorders of the skin such as allergic contact dermatitis,
 CC psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
 CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
 CC sequence represents human LST-2.
 CC
 XX
 XX Sequence 380 AA:
 SQ
 Query Match 98.5%; Score 2008; DB 23; Length 380;
 Best Local Similarity 100.0%; Pred. No. 1.8e-202;
 Matches 380; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 247 QVICSOLEIKTIOISLPKALOERYLLVREYEDLAPAPVAQTSRMTEFVGLFPLPHIQTV 306
 DB 241 QVICSOLEIKTIOISLPKALOERYLLVREYEDLAPAPVAQTSRMTEFVGLFPLPHIQTV 300
 QY 307 HNITRGKMGDHAFTNARDALNVSAWRMSLPYERVSRLQKACGDANMLGRIHVRSEQ 366
 DB 301 HNITRGKMGDHAFTNARDALNVSAWRMSLPYERVSRLQKACGDANMLGRIHVRSEQ 360
 QY 367 EQRNLLDLSTWTWYPEQIH 386
 DB 361 EQRNLLDLSTWTWYPEQIH 380
 RESULT 4
 ID AAY79219 standard; Protein, 386 AA.
 AC AAY79219;
 XX
 XX 19-JUN-2000 (first entry)
 XX DT
 XX DE Human transferrase TRNSFS-11.
 XX
 XX Transferrase; TRNSFS-11; human; antitumor; cell proliferation;
 KW gastrointestinal disorder; developmental disorder;
 KW genetic disorder; neurological disorder; reproductive disorder;
 KW smooth muscle disorder; immunological disorder; inflammation;
 KW diagnosis; therapy; N-acetylglucosamine 6-O-sulfotransferase.
 XX
 XX Homo sapiens.
 XX
 XX
 XX Key Location/Qualifiers
 FT Modified-site 121
 FT /note= "potential O-phosphorylation"
 FT Modified-site 107
 FT /note= "potential O-phosphorylation"
 FT Modified-site 217
 FT /note= "potential O-phosphorylation"
 FT Modified-site 252
 FT /note= "potential O-phosphorylation"
 FT Modified-site 364
 FT /note= "potential O-phosphorylation"
 FT Modified-site 380
 FT /note= "potential O-phosphorylation"
 FT Modified-site 35
 FT /note= "potential O-phosphorylation"
 FT Modified-site 50
 FT /note= "potential O-phosphorylation"
 FT Modified-site 81
 FT /note= "potential O-phosphorylation"
 FT Modified-site 287
 FT /note= "potential O-phosphorylation"
 FT Modified-site 243
 FT /note= "potential O-phosphorylation"
 FT Modified-site 30
 FT /note= "potential N-glycosylation"
 FT Modified-site 308
 FT /note= "potential N-glycosylation"
 FT Modified-site 329
 FT /note= "potential N-glycosylation"
 FT Domain 7..23
 FT /note= "transmembrane domain"
 XX
 XX WO200014251-A2.
 XX PN
 XX 16-MAR-2000.
 XX PD
 XX 09-SEP-1999;
 XX PE 99WO-US20989.
 XX 10-SEP-1998;
 XX PR 98US-0150657.
 XX 04-NOV-1998;
 XX PR 98US-0186779.
 XX 11-MAY-1999;
 XX PR 99US-0133642.

XX (INCY-) INCYTE PHARM INC.
 XX Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;
 PI Hillman JL, Azimzal Y;
 XX
 DR MPI: 2000-256996/22.
 DR N-PSDB: AA294211.
 XX
 XX Human transferase proteins useful for preventing, diagnosing and
 PT treating cancers and developmental, gastrointestinal, genetic,
 PT immunological, neurological, reproductive and smooth muscle disorders -
 XX
 PS Claim 1; Page 90-91; 113pp; English.
 XX
 CC The present sequence is that of human transferase TRNSFS-11, 1 of
 CC 15 claimed novel human transferase proteins of the invention (see
 CC AA979209-23). The sequence was deduced from a cDNA clone (see
 CC AA294211) isolated from a galbladder library. It shows homology to
 CC mouse N-acetylglucosamine 6-O-sulfotransferase. TRNSFS-11 is
 CC expressed in dermal and gastrointestinal tissues,
 CC especially those associated with inflammation and cell
 CC proliferation. The new human transferases and polynucleotides can
 CC be used in the diagnosis, prevention and treatment of cancer,
 CC developmental disorders, gastrointestinal disorders, genetic
 CC disorders, immunological disorders, neurological disorders,
 CC reproductive disorders, and smooth muscle disorders. The
 CC polypeptides can also be used to raise antibodies, and to screen
 CC for agonists and antagonists of transferase activity.
 CC
 SQ Sequence 386 AA:
 Query Match 95.0%; Score 1936; DB 21; Length 386;
 Best Local Similarity 93.6%; Pred. No. 7, 2e-195;
 Matches 369; Conservative 3; Mismatches 14; Indels 0; Gaps 0;
 QY 1 MLTPKMKLLFLVSOMATLALFFHMYSHNLSLSKKAQPERHNVLYLSMRGSSFWG 60
 DB 1 MLTPKMKLLFLVSOMATLALFFHMYSHNLSLSKKAQPERHNVLYLSMRGSSFWG 60
 QY 61 LFGQHPDVFLMEPRMHWMTFKOSTAMMLHMAVNDLIRAVFLCDMSVFDAYMEGPRRQ 120
 DB 61 LFGQHPDVFLMEPRMHWMTFKOSTAMMLHMAVNDLIRAVFLCDMSVFDAYMEGPRRQ 120
 QY 121 SSLSFQWENSRALCSAPACDIIPODETTPRAHCRILCSQPFVEYVEKACSYSHVYLKEVR 180
 DB 121 SSLSFQWENSRALCSAPACDIIPODETTPRAHCRILCSQPFVEYVEKACSYSHVYLKEVR 180
 QY 181 FENLQSLVPLKDPSTLNLHIVLRDPRAVFRSRETKGDLMDISRTYMGQHEQKLKED 240
 DB 181 FENLQSLVPLKDPSTLNLHIVLRDPRAVFRSRETKGDLMDISRTYMGQHEQKLKED 240
 QY 241 QPYVYMOVICOSOLEYKTIQSLPKALQERYLLVREDELARAPVACTSRMYEFGLEFLP 300
 DB 241 QPYVYMOVICOSOLEYKTIQSLPKALQERYLLVREDELARAPVACTSRMYEFGLEFLP 300
 QY 301 HLOTWVHNITRKGMDHAFHTNARDALNVSQAMRSLPEYKYSRLQKACGDMNLGYR 360
 DB 301 HLOTWVHNITRKGMDHAFHTNARDALNVSQAMRSLPEYKYSRLQKACGDMNLGYR 360
 QY 361 HVRSQEQORNLLDLSTWTVPQI 386
 DB 361 HVRSQEQORNLLDLSTWTVPQI 386
 RESULT 5
 AA939919
 ID AA939919 standard; Protein; 388 AA.
 AC AA939919;
 XX
 XX 08-DEC-1999 (first entry)
 DT
 XX

DE Mouse glycosyl sulfotransferase-3 protein sequence.
 XX
 XX Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
 KW selectin binding interaction; Inflammation; lymphocyte homing; mouse;
 KW secondary lymph organ.
 XX
 XX Mus sp.
 XX
 XX MO9949018-A1.
 XX
 XX 30-SEP-1999.
 XX
 XX 26-FEB-1999; 99MO-0504316.
 XX
 XX 20-MAR-1998; 98US-0045284.
 XX 12-NOV-1998; 98US-0190911.
 XX
 PA (REGC) UNIV CALIFORNTA.
 PA (SYNT) SYNTX USA INC.
 PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
 DR MPI: 1999-580442/49.
 DR N-PSDB: AA20793.
 XX
 XX Human and murine glycosyl sulfotransferase 3 and related
 PT polynucleotides
 PS Claim 2; Fig 4; 59pp; English.
 XX
 CC This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.
 CC
 SQ Sequence 388 AA:
 Query Match 74.3%; Score 1513.5; DB 20; Length 388;
 Best Local Similarity 72.8%; Pred. No. 2, 2e-150;
 Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;
 QY 1 MLTPKMKLLFLVSOMATLALFFHMYSHNLSLSKKAQPER-MHYLYLSMRGSSFWG 59
 DB 1 MLTPKMKLLFLVSOMATLALFFHMYSHNLSLSKKAQPER-MHYLYLSMRGSSFWG 58
 QY 60 QLEGOHPDVFLMEPRMHWMTFKOSTAMMLHMAVNDLIRAVFLCDMSVFDAYMEGPRR 119
 DB 60 QLEGOHPDVFLMEPRMHWMTFKOSTAMMLHMAVNDLIRAVFLCDMSVFDAYMEGPRR 118
 QY 59 QLEGOHPDVFLMEPRMHWMTFKOSTAMMLHMAVNDLIRAVFLCDMSVFDAYMEGPRR 118
 DB 59 QLEGOHPDVFLMEPRMHWMTFKOSTAMMLHMAVNDLIRAVFLCDMSVFDAYMEGPRR 118
 QY 120 QSSLSFQWENSRALCSAPACDIIPODETTPRAHCRILCSQPFVEYVEKACSYSHVYLKEV 179
 DB 120 QSSLSFQWENSRALCSAPACDIIPODETTPRAHCRILCSQPFVEYVEKACSYSHVYLKEV 178
 QY 179 QSSLSFQWENSRALCSAPACDIIPODETTPRAHCRILCSQPFVEYVEKACSYSHVYLKEV 178
 DB 179 QSSLSFQWENSRALCSAPACDIIPODETTPRAHCRILCSQPFVEYVEKACSYSHVYLKEV 178
 QY 180 RPNLQSLVPLKDPSTLNLHIVLRDPRAVFRSRETKGDLMDISRTYMGQHEQKLKED 239
 DB 180 RPNLQSLVPLKDPSTLNLHIVLRDPRAVFRSRETKGDLMDISRTYMGQHEQKLKED 238
 QY 239 RPNLQSLVPLKDPSTLNLHIVLRDPRAVFRSRETKGDLMDISRTYMGQHEQKLKED 238
 DB 239 RPNLQSLVPLKDPSTLNLHIVLRDPRAVFRSRETKGDLMDISRTYMGQHEQKLKED 238
 QY 240 DQPYVYMOVICOSOLEYKTIQSLPKALQERYLLVREDELARAPVACTSRMYEFGLEFL 299
 DB 240 DQPYVYMOVICOSOLEYKTIQSLPKALQERYLLVREDELARAPVACTSRMYEFGLEFL 298
 QY 299 DQPYVYMOVICOSOLEYKTIQSLPKALQERYLLVREDELARAPVACTSRMYEFGLEFL 298
 DB 299 DQPYVYMOVICOSOLEYKTIQSLPKALQERYLLVREDELARAPVACTSRMYEFGLEFL 298
 QY 300 PHLOTWVHNITRKGMDHAFHTNARDALNVSQAMRSLPEYKYSRLQKACGDMNLGYR 359
 DB 300 PHLOTWVHNITRKGMDHAFHTNARDALNVSQAMRSLPEYKYSRLQKACGDMNLGYR 358
 QY 359 PHLOTWVHNITRKGMDHAFHTNARDALNVSQAMRSLPEYKYSRLQKACGDMNLGYR 358
 DB 359 PHLOTWVHNITRKGMDHAFHTNARDALNVSQAMRSLPEYKYSRLQKACGDMNLGYR 358
 QY 360 HVRSQEQORNLLDLSTWTVPQI 385
 DB 360 HVRSQEQORNLLDLSTWTVPQI 384

RESULT 6
 AAB41947
 ID AAB41947 standard; Protein: 418 AA.
 AC AAB41947;
 DT 08-FEB-2001 (first entry)
 DE Human ORFX ORF1711 polypeptide sequence SEQ ID NO:3422.
 XX
 KW Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
 KW anticonvulsant; osteopathic; antihypertensive; immunosuppressive; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antineoplastic; antihypertensive;
 KW antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 XX MO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 9905-0127607.
 PR 02-APR-1999; 9905-0127636.
 PR 05-APR-1999; 9905-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 PA
 PI Shinkets RA, Leach M;
 XX
 DR N-PSDB; AAC76156.
 DR
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 XX
 PS Claim 11; Page 2599-2600; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparasitic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antihypertensive; immunosuppressive;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antineoplastic;
 CC antihypertensive; antineoplastic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX
 SO Sequence 418 AA;
 Query Match 59.2%; Score 1205.5; DB 21; Length 418;
 Best Local Similarity 61.2%; Pred. No. 6.6e-118;
 Matches 241; Conservative 47; Mismatches 79; Indels 27; Gaps 6;
 QY 1 MLFPKKMLLFLVSOMAILALFFHMYSHNTSSLSMKRQPRMHVVLSSRSGSPFGQ 60
 DB 41 LLLAQTCCLLFLFISRP-----GSSPAGGEDRVHVLSSSRSGSPFGQ 86
 QY 61 LFGQHPVFLYMEPAMHMYMTFKQSTAMMLHAAVRLDILRAVLCDSYFDAYMEPGPRQ 120
 DB 87 LFSQHPVFLYMEPAMHMYMTLSSGSAATLHAAVRLDILRAVLCDSYFDAYMEPGPRQ 146
 QY 121 SLPFWENSRAALCAPACDIIIPDEIIPRAHCRLLCSQOPEVEYKACRSYSHVLYEVR 180
 DB 147 SLPFWENSRAALCAPACDIIIPDEIIPRAHCRLLCSQOPEVEYKACRSYSHVLYEVR 206
 QY 181 FPNQSLYPLKDSNLHIVHVRDPAVFRSPERTKGDLMDSRIYMGHDEKLRKD 240
 DB 207 FPNQSLYPLKDSNLHIVHVRDPAVFRSPERTKGDLMDSRIYMGHDEKLRKD 265
 QY 241 QPYVMQVICSQOLEIYK--TIQSLPKALQERYLLVREYEDLARA PVATSHMEFVGLIEF 298
 DB 266 PHLLIREVCRSHRIAEATLKP--PPLRGRYRLVREYEDLARA PVATSHMEFVGLIEF 324
 QY 299 LPHLQTVHNTTRKGMGD--HAFHTNARDALNVSQAWRMSLYEKYSRLQACGDAMNL 356
 DB 325 TPQLEAMIHNTTSGGIGKPIEAFTSSRNARNVSQAWRHALPFTKILRQVEGACALQL 384
 QY 357 LGRYRVSDEQRNLDDL-----STWVPE 383
 DB 385 LGRYRVSADQQRDLTDLVLRGPDHPSWASPD 418
 XX
 RESULT 7
 ID ABB81557 standard; Protein: 418 AA.
 AC ABB81557;
 DT 05-SEP-2002 (first entry)
 XX
 DE Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.
 KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological.
 XX
 OS Mus musculus.
 XX
 PN US2002061562-A1.
 XX
 PD 23-MAY-2002.
 XX
 PF 09-AUG-2001; 2001US-0927602.
 XX
 PR 11-AUG-2000; 2000US-325773P.
 PA (FUKU/) FUKUDA M N.
 PA (AKAW/) AKAWA T O.
 PI Fukuda MN, Akawa TO;
 XX
 DR WPI: 2002-507643/54.
 XX
 CC New nucleic acid encoding corneal
 CC N-acetylglucosamine-6-sulfotransferase, useful for treatment,
 CC monitoring and diagnosis of macular corneal dystrophy -
 PT
 XX Example 5; Page 24-25; 69pp; English.
 PS

CC The present invention describes human corneal
 CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze
 CC sulfation of keratan sulfate (KS). Also described is a method for
 CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
 CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
 CC and has ophthalmological activity. (I) can be used to treat or prevent
 CC macular corneal dystrophy types I or II. (I) makes possible treatment
 CC of MCD without requiring keratoplasty or keratectomy. The present
 CC sequence represents mouse intestinal N-acetylglucosamine-6-
 CC sulfotransferase, which is given in comparison with (I) in the
 CC exemplification of the present invention.

SO Sequence 418 AA;

Query Match 50.4%; Score 1028; DB 23; Length 418;
 Best Local Similarity 56.6%; Pred. No. 3.2e-99;
 Matches 215; Conservative 43; Mismatches 112; Indels 10; Gaps 6;

QY 1 MLTPK--KMKLLFLYSQAMLLALFFHMTSHNLSLSMKAPRMVTVLSSRRSSSY 58
 DB 25 MLPRSSSTVMSLMLVQGLVFE--LVSRQVPS--SPAGLGRVHVLSSRRSSSY 80
 QY 59 GOLFGQHPDVFYLMPEAMHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAVMEPGPR 118
 DB 81 GOLFGQHPDVFYLMPEAMHVMWTLSSGSAFALHMAVRDLIRSVFLCDMDVFDAYL-PMRR 139
 QY 119 ROSSLFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPEVEVEKACRSYSHVYLKE 178
 DB 140 NISDLFQWAVSRALCSPPVCEAFARCNISSEVECKFLCATRPGLAQECSSYSHVYLKE 199
 QY 179 VRFENQSLYPLIKDPSLMLHTVLRDPRVFRSEKRGDMLDSRYMGHEKLLK 238
 DB 200 VRFENQSLYPLISDPALMLRIVHLRDPRAVLRSEQFAKALARNGVLTNGTWV-E 258
 QY 239 EDQPYVMOVICOSQLEIKYKTIQSLPKAQERYLTVREEDLARAQAQTSRMTEFEVLEF 298
 DB 259 ADPRLVAVNEVCGRSHVRIEALHKKPPFLODRRLVRYEDLARDPLTVIRELTAFTGLGL 318
 QY 299 LPHLOTVNHNITRGKMG--DHAFTNARDALNVSQAMWSLPEKVSRLQKACGDAMNL 356
 DB 319 TPOLOTVNHNITRGSGPARREAFKTSRDALSVQAMWHTLPFAKIRVQELCGGALQL 378
 QY 357 LGRYHVSSEQORNLIDL 376
 DB 379 LGRYHVSSEQORNLIDL 398

RESULT 8

AAV72640
 ID AAV72640 standard; Protein; 395 AA.

AC AAV72640;
 DT 02-MAY-2001 (first entry)

DE Human glycosyl sulfotransferase-4beta (GST-4beta).

XX Human; glycosyl sulfotransferase-4beta; GST-4beta; immunosuppressive;
 XX therapy; selectin binding inhibitor; gene therapy; inflammation;
 XX systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 XX glomerulonephritis; nodosa; polymyositis; systemic sclerosis; dermatitis;
 XX Hashimoto's disease; Grave's disease; hypoparathyroidism; anemia;
 XX demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 XX myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 XX asthma; hypersensitivity; rheumatic fever; tissue rejection;
 XX chromosome 16q23.1.

OS Homo sapiens.
 XX WO200106015-A1.
 PN 25-JAN-2001.
 XX

XX 19-JUL-2000; 2000MO-0519741.
 XX 20-JUL-1999; 99US-0144694.
 PR 13-JUL-2000; 2000US-0593828.
 XX (REGC) UNIV CALIFORNIA.
 XX Rosen SD, Lee JK, Hemmerich S;
 DR WPI: 2001-138471/14.
 DR N-PSDB; AAD02697, AAD02700.
 XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 PT diagnostic and therapeutic agent screening applications
 PS claim 3; Fig 4B; 128pp; English.

CC The present sequence is human glycosyl sulfotransferase-4beta (GST-4
 CC beta). GST-4 gene is found on chromosome 16q23.1.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand. GST and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
 CC disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
 CC anemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.

SO Sequence 395 AA;

Query Match 50.0%; Score 1019; DB 22; Length 395;
 Best Local Similarity 54.5%; Pred. No. 2.6e-98;
 Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY 2 LTPKMKLLFLYSQAMLLALFFHMTSHNLSLSMKAPRMVTVLSSRRSSSYVQGL 61
 DB 14 LLLAQFLLFLFVSRP-----GPSSPAGGEARVHVLSSRRSSSYVQGL 59
 QY 62 FGOHPDVFYLMPEAMHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAVMEPGPR 121
 DB 60 FNOHPDVFYLMPEAMHVMWTLSSGSAATLHMAVRDLIRSVFLCDMDVFDAYL-PMRRNLS 118
 QY 122 SLFQWENSRLCSAPACDIIPODEIIPRAHCRLLCSQOPEVEVEKACRSYSHVYLKEVR 181
 DB 119 DLFQWAVSRALCSPPVCEAFARCNISSEVCKFLCATRPGFLAREKCRSYSHVYLKEVR 178
 QY 182 FNIQSLYPLIKDPSLMLHTVLRDPRVFRSEKRGDMLDSRYMGHEKLLKEQ 241
 DB 179 FNIQSLYPLISDPALMLRIVHLRDPRAVLRSEQFAKALARNGVLTNGTWV-EADP 237
 QY 242 PYVMOVICOSQLEIKYKTIQSLPKAQERYLTVREEDLARAQAQTSRMTEFEVLEFL 299
 DB 238 GLRVAVNEVCGRSHVRIEALHKKP-PPFLGRRLVRYEDLARDPLTVIRELTAFTGLSLT 296
 QY 300 PHLOTVNHNITRGKMG--DHAFTNARDALNVSQAMWSLPEKVSRLQKACGDAMNL 357
 DB 297 PQLFAMHNITRGSGPARREAFKTSRRNALNVSQAMWHTLPFAKIRVQELCGGALQL 356
 QY 358 GYRHVSSEQORNLIDL-----STVTPVQIH 386
 DB 357 GYRPVYSSEQORNLIDLVLPRGLNGFTWASSTASH 392

RESULT 9

ABB81554

ID ABB81554 standard; Protein: 395 AA.
 AC ABB81554:
 XX
 DT 05-SEP-2002 (first entry)
 XX
 DE Human corneal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:2.
 XX
 KW Human: N-acetylglucosamine-6-sulfotransferase; enzyme: GICNAC6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological; chromosome 16q22.
 XX
 OS Homo sapiens.
 PN US2002061562-A1.
 XX
 PD 23-MAY-2002.
 XX
 PF 09-AUG-2001; 2001US-0927602.
 XX
 PR 11-AUG-2000; 2000US-325773P.
 XX
 PA (FUKU/) FUKUDA M N.
 PA (AKAM/) AKAMA T O.
 XX
 PI Fukuda MN, Akama TO;
 XX
 DR WPI: 2002-507643/54.
 DR N-PSDB: ABN89506.
 XX
 PT New nucleic acid encoding corneal
 PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,
 PT monitoring and diagnosis of macular corneal dystrophy -
 XX
 PS Claim 13; Fig 1A-D; 69pp; English.
 CC
 CC The present sequence represents human corneal
 CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyse
 CC sulfation of keratan sulfate (KS). Also described is a method for
 CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
 CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
 CC and has ophthalmological activity. (I) can be used to treat or prevent
 CC macular corneal dystrophy types I or II. (I) makes possible treatment
 CC of MCD without requiring keratinoplasty or keratectomy.
 CC
 XX
 SO Sequence 395 AA:
 Query Match 50.0%; Score 1019; DB 23; Length 395;
 Best Local Similarity 54.5%; Pred. No. 2,66-98;
 Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

DB 297 POLEAMHNTHGSGPGARREAFKTSRRNALNSQAWRHLPFAKIRHVOELCAGALQL 356
 QY 358 GYRVRSQEQEORNLIDL-----STWYVPEQIH 386
 DB 357 GYRPYSDEQRNALDLVLPRLGNGFTWASSTASH 392
 RESULT 10
 ID AAE15438
 ID AAE15438 standard; Protein: 395 AA.
 XX
 AC AAE15438;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Human drug metabolizing enzyme (DME)-5.
 KW Human; drug metabolizing enzyme; gene therapy; autoimmune disorder;
 KW inflammatory disorder; acquired immune deficiency syndrome; infection;
 KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
 KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
 KW cancer; endocrine disorder; hypothyroidism disorder; pituitary disorder;
 KW gastrointestinal disorder; metabolic disorder; developmental disorder;
 KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
 KW goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
 KW DME-5.
 XX
 OS Homo sapiens.
 XX
 XX Key Location/Qualifiers
 XX Peptide 1..32
 XX /label- Signal_peptide
 XX Protein 33..395
 XX /note- "Human mature DME-5 protein"
 XX Peptide 1..35
 XX /label- Signal_peptide
 XX Protein 36..395
 XX /note- "Human mature DME-5 protein"
 XX
 XX WO200179468-A2.
 XX
 XX 25-OCT-2001.
 PD
 PF 12-APR-2001; 2001WO-US11869.
 XX
 PR 13-APR-2000; 2000US-197590P.
 PR 19-APR-2000; 2000US-198403P.
 PR 28-APR-2000; 2000US-200185P.
 PR 05-MAY-2000; 2000US-202234P.
 PR 11-MAY-2000; 2000US-203509P.
 XX
 XX (INCY-) INCYTE GENOMICS INC.
 XX
 XX Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;
 XX Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;
 XX Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;
 XX Au-Young J;
 XX
 DR WPI: 2002-066363/09.
 DR N-PSDB: AAD24670.
 XX
 PT Novel isolated human drug metabolizing enzymes referred as DME 1-10,
 PT useful for diagnosing, treating, or preventing disorders associated
 PT with aberrant expression of DME such as allergy, anemia, asthma,
 PT infertility -
 XX
 PS Claim 1a; Page 131-132; 143pp; English.
 CC The invention relates to human drug metabolizing enzymes referred as
 CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides
 CC of the invention are useful for assessing toxicity of test compounds
 CC and in gene therapy. Sequences of the invention are useful in the
 CC diagnosis, prevention and treatment of autoimmune/inflammatory

OY 298 FLEHDTWVNTIRGSGM - DHAPTNARALNVSQAWRSLEYEYSRLQKAGSDAMN 353S
| | | | | | | | | | | | | | | | | | | | | : | : | : | : | :
DB 295 LTFOLDTWNHTIHGGSPPEARREARTKTSRDALSQAWMRHTLPANIKRRDLCGGALO 354S

OY 356 LLGYRHRVSRFQEQNRLLDDL 376
| | | | | | | | | | | | | | | | | | | | | : | : | : | : | :
DB 355 LLGYRHSVSELRDRDLSDL 375

RESULT 12
ID AA011275
AA011275 standard; Protein; 395 AA.
AC AA011275;
XX
DE 12-MAR-2002 (first entry)
XX
MU Murine intestinal-GlcNAc 6-sulfotransferase (I-GlcNAc6ST) protein.
KM Mouse; beta1,3gnt; beta1,3-N-acetylglucosaminyltransferase; MECA-79;
KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
KW allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
KW delayed-type hypersensitivity reaction; hyperplastic thymus; antileuker;
KW antiinflammatory; antipsoriatic; antidiabetic; dermatological;
XX antiallergic; intestinal-GlcNAc 6-sulfotransferase; I-GlcNAc6ST.
OS Mus musculus.
XX
PN WO200185177-A1.
XX
PD 15-NOV-2001.
XX
PE 10-MAY-2001; 2001MO-US15452.
XX
PR 11-MAY-2000; 2000US-0569320.
PA (BURN-) BURNHAM INST.
XX
PI Fukuda M, Yeh J, Haraoka N:
DR WPI; 2002-075226/10.
DR N-PSDB; AAS16948.
XX
PT New enzyme, useful for modifying acceptor molecule, comprises an
PT isolated L-selectin sulfotransferase-2 that directs expression of
PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
PT intestinal GlcNAc 6-sulfotransferase
PS
PS Claim 28; Fig 10; 98pp; English.
XX
XX The present invention provides a method of modifying an acceptor moleculat
CC by contacting the acceptor with an isolated
CC beta1,3-N-acetylglucosaminyltransferase (beta1,3gnt) or an active
CC fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The
CC invention also provides a method of treating or preventing an
CC L-selectin-mediated condition by reducing the expression or activity of a
CC beta1,3gnt that directs expression of a MECA-79 antigen. This can be done
CC by administering to the subject an oligosaccharide L-selectin antagonist
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
CC administering antibody material that specifically binds beta1,3gnt,
CC and/or a beta1,3gnt antisense nucleic acid molecule. L-selectin
CC sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
CC Alternatively, the expression or activity of LST-2 or its active
CC fragment can be reduced in combination with reducing the expression or
CC activity of beta1,3gnt. The method is useful for treating L-selectin
CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis,
CC psoriasis and lichen planus, lymphomas, chronic pneumonitis, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC sequence represents mouse I-GlcNAc6ST..
XX

Seq	Sequence	395 AA:	49.9%:	Score 1017.5:	DB 23:	Length 395:
Query Match			56.4%:	Pred. No. 3.8e-98:		
Best Local Similarity			43:	Mismatches 112:	Indels 11:	Gaps 7
Matches 215:						
OY	1	MLLPK--KMKLLLEFVSQMAITLALFPHMYSHNISSISMKAQPERMHVTLVLSMRSGSSPV	58			
DB	1	MRLPFESTVMLSTLMVGTGLV--LVSRQVPS--SPAGLGERVHVLVLSMRSGSSPV	56			
OY	59	GOLFQGHDFVEFLMEPAHWVMTFKQSTAMLMHMAVRDLIRAVFLCDMSFPDAMEPGR	118			
DB	57	GOLFQGHDFVEFLMEPAHWVMTLQSGSAPALHMAVRDLIRSVFLCDMDVFDAYL--PMRR	115			
OY	119	ROSSLEQWENSRALCSAPACDIIPODEIIPRAHCRLLCSQOPPEVEYKCRSYHYLKE	178			
DB	116	NISDFQWAVSFLCSPVCEAFARGNISSEVCKPLCATRPFGLAQCASSSHVYLKE	175			
OY	179	VREFLQSLYPLLPKPSLNLTHVHVPDRAVFRSREPRKGLMIDSRVGMQHEKLRK	238			
DB	176	VREFLQSLYPLLPSPALNLRTVHVRDRAVLRSEQYAKALARNGIVLTNGTW--E	234			
OY	239	EDQPRYVQVLCQSOLSEIYK--TIQSLPKALQERVLRYEDLARAFAVQTSRMVEFGLE	297			
DB	235	ADPRLRVNVENCRRSHVRIAEALNHPRPPLDQRYLRVREDLARDLTYIRRELYAFTGLG	294			
OY	298	FLPHLQTVHVNITRKGKMG--DHAFHTNARDALNVSQAWRSLPYEKVSRLQACGDAMN	355			
DB	295	LTPOLQTVIHNTHTSGPGARREAREKTSRDALSVQAWRHVLPFAKIRRVQELCGALQ	354			
OY	356	LLGYRHVRSQEQRLMLDDL	376			
DB	355	LLGYRHVHSELEQRDLSDL	375			
RESULT 13						
ABB81555						
ID	ABB81555	standard: Protein: 395 AA.				
XX	ABB81555:					
AC						
XX						
DT	05-SEP-2002	(first entry)				
DE		Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3.				
XX						
XX						
KW		Human: N-acetylglucosamine-6-sulfotransferase; enzyme; GLENA6ST;				
KM		corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;				
XX		ophthalmological.				
XX						
OS		Homo sapiens.				
OS		Mus musculus.				
OS		Synthetic.				
XX						
FT	Key	Location/Qualifiers				
FT	Misc-difference	10 /label- Ala, Thr, Val				
FT	Misc-difference	13 /label- Ala, Val, Ser				
FT	Misc-difference	20 /label- Phe, Cys, Gly				
FT	Misc-difference	39 /label- Ala, Asp, Glu				
FT	Misc-difference	96 /label- Val, Met, Ile				
FT	Misc-difference	142 /label- Ala, Thr, Asn				
FT	Misc-difference	147 /label- Ala, Asp, Glu				
FT	Misc-difference	159 /label- Thr, Ser, Gly				
FT	Misc-difference	238 /label- Gly, His, Arg				
FT	Misc-difference	294				

```

FT      /label- Ser, Thr, Gly
FT      M1sc-difference 371
FT      /label- Ala, Thr, Ser
FT      M1sc-difference 380
FT      /label- Leu, Pro, Met
FT      M1sc-difference 382
FT      /label- Gly, His, Ser
FT      M1sc-difference 384
FT      /label- Thr, Ser, Lys
FT      M1sc-difference 390
FT      /label- Ala, Glu
FT      M1sc-difference 391
FT      /label- Ser, Lys
FT      M1sc-difference 392
FT      /label- His, Glu
FT      M1sc-difference 394
FT      /label- Arg, Glu
FT      M1sc-difference 395
FT      /label- Asn, Ser
FT
FT      US2002061562-A1.
FT
FT      23-MAY-2002.
FT
FT      09-AUG-2001; 2001US-0927602.
FT
FT      11-AUG-2000; 2000US-325773P.
FT
FT      (FUKUDA M N.
FT      (AKAMA T O.
FT
FT      Fukuda MN, Akama TO;
FT
FT      WPI; 2002-507643/54.
FT
FT      New nucleic acid encoding corneal
FT      N-acetylglucosamine-6-sulfotransferase, useful for treatment,
FT      monitoring and diagnosis of macular corneal dystrophy
FT
FT      Example 5; Fig 2A-B; 69pp; English.
FT
FT      The present invention describes human corneal
FT      N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze
FT      sulfation of keratan sulfate (KS). Also described is a method for
FT      monitoring the effect of treatments for macular corneal dystrophy (MCD),
FT      and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
FT      and has ophthalmological activity. (I) can be used to treat or prevent
FT      macular corneal dystrophy types I or II. (I) makes possible treatment
FT      of MCD without requiring keratoplasty or keratectomy. The present
FT      sequence represents a consensus N-acetylglucosamine-6-sulfotransferase
FT      which is given in the exemplification of the present invention.
FT
FT      Sequence 395 AA:
FT
FT      Query Match 49 7%; Score 1013.5; DB 23; Length 395;
FT      Best Local Similarity 55.9%; Pred. No. 1e-97;
FT      Matches 212; Conservative 45; Mismatches 101; Indels 21; Gaps 6;
FT
FT      2 LTPKKMKLLFLVYQMAIIALFFHMYSHNIISSLSKMAOPERMHVILSSMRSGSSFVGDL 61
FT      14 LLAAQYXLLFLVYSR-----GSSPAGGEXRNVIVLSSMRSGSSFVGDL 59
FT
FT      62 FQOHDPVFLMEPAHVMFTFKOSTAMMLMAVRDILIRAVFLCDMSVFADYMEGPPRRQS 121
FT      60 FQOHDPVFLMEPAHVMFTTLLSQGSNAATLHMAVRDILIRAVFLCDMDVFDAYL-PMRRNDS 118
FT
FT      122 SLFQWNSRALCSAPRCDIIPQDELIPRAHCRILCSQPFVEVERKCRSYSHVYLKEVRF 181
FT      119 DLFQWNSRALCSAPRCASFPKXISSEXYCKPLCARQPFYLAREACRSYSHVYLKEVRF 178
FT
FT      182 FNLQSLYPLKDPSSLHLIVHLVDRPRAVRSRERTKGLMIDSRIVMGQHEQKLKEQD 241
FT      179 FNLQSLYPLKDPSSLHLIVHLVDRPRAVRSRERTKGLMIDSRIVMGQHEQKLKEQD 241

```

```

OY      242 PYVMQVICOSELEIK--TIOSLPKALOERYLVREYEDLARAPVQTSRMVEFVGLLEFL 299
OY      238 XLRVAREVCGRSHRIRAEATIKP-PPLFGRRHVLVEFDLAREPLAEIRALVAFGLXLT 296
OY      300 PHLQTVNHNITRCKGNG--DHAFTNARDALNVQAMRWSLPEYKYSRLQKACGDAMNLL 357
OY      297 POLEAMIHNIHTHSGCGARREAFKTSRRNALNVQAMRHALPFAKIRVQELCAGALQTL 356
OY      358 GYRHRVRSBOORLLDLDL 376
OY      357 GYRHYSEDEQRLXLDLV 375
OY
OY      RESULT 14
OY      AAY72639
OY      AAY72639 standard; Protein; 390 AA.
OY
OY      AAY72639;
OY
OY      02-MAY-2001 (first entry)
OY
OY      Human glycosyl sulfotransferase-4alpha (GST-4alpha).
OY
OY      Human: glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
OY      therapy; selectin binding inhibitor; gene therapy; inflammation;
OY      systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
OY      polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
OY      glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;
OY      Hashimoto's disease; Grave's disease; hypoparathyroidism; anemia;
OY      demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
OY      myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
OY      asthma; hypersensitivity; rheumatic fever; tissue rejection;
OY      chromosome 16q23.1.
OY
OY      Homo sapiens.
OY
OY      WO200106015-A1.
OY
OY      25-JAN-2001.
OY
OY      19-JUL-2000; 2000MO-US19741.
OY
OY      20-JUL-1999; 99US-0144694.
OY      13-JUL-2000; 2000US-0593828.
OY
OY      (REGC ) UNIV CALIFORNIA.
OY
OY      Rosen SD, Lee JK, Hemmerich S;
OY
OY      WPI; 2001-138471/14.
OY      N-PSDB; AAD02697, AAD02698, AAD02699.
OY
OY      New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
OY      diagnostic and therapeutic agent screening applications
OY
OY      Claim 3; Fig 1; 128pp; English.
OY
OY      The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
OY      alpha). GST-4 gene is found on chromosome 16q23.1.
OY      GST is a type 2 membrane protein useful for inhibiting a binding event
OY      between a selectin and a selectin ligand, which comprises contacting the
OY      selectin with a non-sulphated selectin ligand, GST and a small molecular
OY      agent that inhibits the sulphation activity of GST. GST is also useful
OY      in inhibiting a selectin mediated binding event. GST is useful in gene
OY      therapy to treat disorders such as acute or chronic inflammation.
OY      systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
OY      nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
OY      glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
OY      disease, Grave's disease, adrenailitis, hypoparathyroidism, pernicious
OY      anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
OY      dermatitis, myocarditis, regional enteritis, adult respiratory distress
OY      syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,

```

CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC during transplantation.
CC Note: The present sequence is also shown in sequence listing (page
CC no: 56) but lacks four nucleotides at its 3' end.
XX
SQ Sequence 390 AA:
Query Match 49.5%; Score 1008; DB 22; Length 390;
Best Local Similarity 52.3%; Pred. No. 3.7e-97;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;
QY 1 MLPRKKMLLFLVSOAIIALFFHMYSHNISLSMKRQPRMHVLYLSWRSSSFYQ 60
DB 14 LLLAQTTCLLFIISRP-----GPSPPAGGEDRVHVLVLSWRSSSFYQ 59
QY 61 LFGQHPDVFYLMERPAHMYMTFKOSTAMLMNAVRLIRAVFLCDMSYFDAYMEGPRQ 120
DB 60 LFSQHPDVFYLMERPAHMYMTFKOSTAMLMNAVRLIRAVFLCDMSYFDAYM-PQSHNL 118
QY 121 SSLFQWENSRLACSAFACDIIPODEIIPRAHCRLLCSQOPEVEVEKACRSYSHVLYKEVR 180
DB 119 SAFNMTATSRALCSPPACSAFPRGTISKQDYCKTLCITQPSLARACRSYSHVLYKEVR 178
QY 181 FFNIQSLYPLKDSNLNLIYHLYVDPRAVRSRERTKGMIDSRIVMGQEQKLKED 240
DB 179 FFNIQSLYPLKDSNLNLIYHLYVDPRAVRSRERTKGMIDSRIVMGQEQKLKED 237
QY 241 QPYVMOVIGOSOLEIK--TIQSLPKALOERYLLVREDIARAPVQTSMYEFGIEF 298
DB 238 PHLRLIREVCRSHVRIAEATLKP-PPLRGRIYLVREDIARAPVQTSMYEFGIEF 296
QY 299 LPHLOTWVHNITRGKMGD--HAFHTNARDALNVSQAWMSLPEKYSRLQKAGDAMNL 356
DB 297 TPQLEAMWHTHTSGICKPIEAFHTSSRNARNVSQAWRHALLPFTKILRVQEVGAGALQL 356
QY 357 LGYRHVSEQQRNLLDIL-----STWTVPE 383
DB 357 LGYRHVSEQQRNLLDIL-----STWTVPE 390
RESULT 15
ABB81556
ID ABB81556 standard; Protein; 390 AA.
XX
AC ABB81556;
XX
DT 05-SEP-2002 (first entry)
XX
DE Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.
XX
KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GICNAC6ST;
KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
KW ophthalmological.
XX
OS Homo sapiens.
XX
PN US2002061562-A1.
XX
PD 23-MAY-2002.
XX
PE 09-AUG-2001; 2001US-0927602.
XX
PR 11-AUG-2000; 2000US-325773P.
XX
PA (FUKU/) FUKUDA M N.
PA (AKAM/) AKAMA T O.
XX
PI Fukuda MN, Akama TO;
XX
DR WPI; 2002-507643/54.
XX
PT New nucleic acid encoding corneal
N-acetylglucosamine-6-sulfotransferase, useful for treatment.

PT monitoring and diagnosis of macular corneal dystrophy
XX
PS Example 5; Fig 2A-B; 69pp; English:
XX
CC The present invention describes human corneal
CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze
CC sulfation of keratan sulfate (KS). Also described is a method for
CC monitoring the effect of treatments for macular corneal dystrophy (MCD),
CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
CC and has ophthalmological activity. (I) can be used to treat or prevent
CC of MCD without requiring keratoplasty or keratectomy. The present
CC sequence represents human intestinal N-acetylglucosamine-6-
CC sulfotransferase, which is given in comparison with (I) in the
CC exemplification of the present invention.
XX
SQ Sequence 390 AA:
Query Match 49.5%; Score 1008; DB 23; Length 390;
Best Local Similarity 52.3%; Pred. No. 3.7e-97;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;
QY 1 MLPRKKMLLFLVSOAIIALFFHMYSHNISLSMKRQPRMHVLYLSWRSSSFYQ 60
DB 14 LLLAQTTCLLFIISRP-----GPSPPAGGEDRVHVLVLSWRSSSFYQ 59
QY 61 LFGQHPDVFYLMERPAHMYMTFKOSTAMLMNAVRLIRAVFLCDMSYFDAYMEGPRQ 120
DB 60 LFSQHPDVFYLMERPAHMYMTFKOSTAMLMNAVRLIRAVFLCDMSYFDAYM-PQSHNL 118
QY 121 SSLFQWENSRLACSAFACDIIPODEIIPRAHCRLLCSQOPEVEVEKACRSYSHVLYKEVR 180
DB 119 SAFNMTATSRALCSPPACSAFPRGTISKQDYCKTLCITQPSLARACRSYSHVLYKEVR 178
QY 181 FFNIQSLYPLKDSNLNLIYHLYVDPRAVRSRERTKGMIDSRIVMGQEQKLKED 240
DB 179 FFNIQSLYPLKDSNLNLIYHLYVDPRAVRSRERTKGMIDSRIVMGQEQKLKED 237
QY 241 QPYVMOVIGOSOLEIK--TIQSLPKALOERYLLVREDIARAPVQTSMYEFGIEF 298
DB 238 PHLRLIREVCRSHVRIAEATLKP-PPLRGRIYLVREDIARAPVQTSMYEFGIEF 296
QY 299 LPHLOTWVHNITRGKMGD--HAFHTNARDALNVSQAWMSLPEKYSRLQKAGDAMNL 356
DB 297 TPQLEAMWHTHTSGICKPIEAFHTSSRNARNVSQAWRHALLPFTKILRVQEVGAGALQL 356
QY 357 LGYRHVSEQQRNLLDIL-----STWTVPE 383
DB 357 LGYRHVSEQQRNLLDIL-----STWTVPE 390

Search completed: August 9, 2003, 16:42:39
Job time : 41.3979 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:40:34 ; Search time 14.9612 Seconds
(without alignments)
2481.152 Million cell updates/sec

Title: US-09-645-078-2

Perfect score: 2038

Sequence: 1 MLPPKMKLLFLVSOMAIL.....EQRNLLDLSTWPEQIH 386

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: PIR_76:**

2: PIR1:**

3: PIR2:**

4: PIR3:**

5: PIR4:**

6: PIR5:**

7: PIR6:**

8: PIR7:**

9: PIR8:**

10: PIR9:**

11: PIR10:**

12: PIR11:**

13: PIR12:**

14: PIR13:**

15: PIR14:**

16: PIR15:**

17: PIR16:**

18: PIR17:**

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	623	30.6	484	2	JC7350
2	606.5	29.8	486	2	JC7351
3	588.5	28.9	484	2	JE0261
4	500.5	24.6	458	2	A57397
5	168	8.2	307	2	E95934
6	116	5.7	388	2	G70729
7	95	4.7	358	2	I40485
8	92.5	4.5	2354	2	T13288
9	92	4.5	359	2	T16350
10	91.5	4.5	338	2	JE0196
11	91	4.5	659	2	S11737
12	91	4.5	1847	2	E64477
13	89.5	4.4	346	2	E86319
14	89.5	4.4	504	2	T07120
15	89.5	4.4	4273	2	C69679
16	89	4.4	524	1	A36205
17	89	4.4	1140	2	G13335
18	89	4.4	1968	1	PN0093
19	88	4.3	359	2	T47187
20	88	4.3	466	2	E84132
21	87.5	4.3	380	2	T42755
22	87.5	4.3	4063	2	T42993
23	87.5	4.3	4101	2	T23630
24	87	4.3	2500	2	G88493
25	86.5	4.2	748	1	S08680
26	86.5	4.2	831	2	S26675
27	86.5	4.2	2352	2	T06077
28	86	4.2	330	2	A40216
29	86	4.2	823	2	S48986

30	86	4.2	869	2	A96558	probable protein k
31	85.5	4.2	344	1	H83928	hypothetical prote
32	85.5	4.2	764	2	S14113	1-phosphatidylino
33	85	4.2	632	2	T18682	hypothetical prote
34	85	4.2	672	2	T34899	probable DNA recom
35	85	4.2	1237	2	T37529	hypothetical prote
36	85	4.2	2688	2	I19477	alpha-A-crystallin
37	84.5	4.1	217	2	J01358	C1 protein - Mica
38	84.5	4.1	646	2	C68683	threonine-tRNA lig
39	84.5	4.1	1031	2	T43458	hypothetical prote
40	84.5	4.1	2493	2	S45734	probable membrane
41	84	4.1	390	2	AB2337	group 2 sigma 70-t
42	84	4.1	1189	2	T42726	guanine nucleotide
43	84	4.1	1217	2	T00270	hypothetical prote
44	84	4.1	1583	2	T14176	probable phosphati
45	83.5	4.1	518	2	E84536	hypothetical prote

ALIGNMENTS

RESULT 1

JC7350 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse

C:Species: Mus musculus (house mouse)

C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000

C:Accession: JC7350

R:Uchimura, K.; Fesakany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, T.

Biochem. Biophys. Res. Commun. 274, 291-296, 2000

A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of

A:Reference number: JC7350; PMID:20374462; PMID:10913333

A:Accession: JC7350

A:Molecule type: mRNA

A:Residues: 1-484 <UCH>

A:Cross-references: DDBJ:AB040710

C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosyl

sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand

C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 30.6% Score 623; DB 2; Length 484;

Best Local Similarity 34.2% Pred. No. 1.5e+45;

Matches 136; Conservative 65; Mismatches 151; Indels 46; Gaps 7;

OY	30	NISSLSKMAQDERHNVIVLSMRSGSSPVGOLFQHPDVFYLMBPAMHMTFKOSTAMM	89
DB	87	NLSAVGEVYDEKQHYVHATWRTGSSSLGELFQHPDVFYLMBPAMHMTFKOSTAMM	146
OY	90	LHMAVRDLIRAVFLCDMSVPDYMPEGRQ-----SSLEWENSRLCSAPACD	139
DB	147	LQALRDLRLSLFQCDPSVLRLVAQPGDGRAPDSANLTYAMLEFRMTNVCSPPLCP	206
OY	140	IIPQDE-----IIPAHCHRLCSQOPFEYEVACGSYSNVLYKEVFFNLSLYPLKDP	195
DB	207	AAPRARADVGLVEKACESTCPVSLRLAEKCRYPVYVIVKIDVLLDGLVPLRLPG	266
OY	196	LNLHIVHLVRDPAVFRSRETKGDLMD-----RIWV-----GONE	233
DB	267	LNLKVVQLFRDPAVFRSRETKGDLMD-----RIWV-----GONE	326
OY	234	QKLKEDQPYV-----VMQVICSQLEIYKTQSLPKALQERYLVREEDLARAVQTSR	289
DB	327	RALPSARADPEFLSALEVICAMLRDLFLRGAPAMLRRLRYLRLREYDLVMOQOARR	386
OY	290	MYEFGVLEFLPHTQWVHNTIRGKMG- DHAFTNARDALVSAWKRSLEFEVYSRLQK	348
DB	387	LIRFSGRITLAALDAFAFMNTRGSAVGDPRFHLISARARAVVWVERLSQEQOVQVET	446
OY	349	ACGDAMNLCGRVHRSBOEORNL-----LDLSTWTPY	381
DB	447	ACAPAMRLAYPRSGDERDKRTVREGTEPLETKANAV	484

RESULT 2

NC7351 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human

C:Species: Homo sapiens (man)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: J07351

R:Uchimura, K.; Passaniti, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, N.; Bloch, M.; Res. Commun. 274, 291-296, 2000

A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a

A:Reference number: J07350; M01D:20374462; PMID:10913333

A:Accession: J07351

A:Molecule type: mRNA

A:Residues: 1-486 <UCH>

A:Cross-references: DDBJ:AB0140711

C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation linkage formation. This enzyme is involved in synthesis of L-selectin ligand in C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 29.8%; Score 606.5; DB 2; Length 486;
Best Local Similarity 37.0%; Pred. No. 3,96-44;
Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;

41 ERHNVLLSSRRSGSSFFVGLGQHPDVEYLMEPAHVMYMTFKSTAMMLHMAVRLIRA 100
100 EKHIVYHATWRTSSSEGLFELNQHDPVEYLFEPMMHLMQALYPGDAESLQALRDLMS 159
101 VFLCDMSVFADYMPG-PRRQ-----SLLQWENSRLCSAPACDITIPDE-----I 146
160 LFRCDVSFLRLYADPPGAAPADPTANTLTALFRMTNRYKICSPPLCPGAPARAEEVGL 219
147 IPRAHCRLLCSQGFEEVEKACRSYSHVLEKEVFNFNLSLPLKPSLNLVHVLVD 206
220 VEDFACERSCPVAIRALAEACRKYPVVYIDVALLDLGLVPLRLRPGNLKVVQLFRD 279
207 PRAVRSRERTGDLMTDS-----RIVM-----GQHEQLKKEPDY 244
280 PRAVHNSRLSKRQGLTRESIDVLTTRRQGRFHRYLLAHGVGAPRGQSRLPAAPRADF 339
245 ---VMQVIOCSOLEIKTIOSLPKALOERYLLVRYDILNAPPAQGSRMTEFVGLFELP 300
340 FLTALTELCAMLRDLFLPARGAPAMLRRLRYLRTEYDVLNPPAQLRLRLFSGLRALA 399
301 HLAQTWNHTKRGKMG-DHAFTNARDALNYSQAMRSLPEKYSRLQKAGDAMNLLGY 359
400 ALDAFALNMTGAYGADRPFHLSARDAREVHAMRRELSREQRYQEAACAPAMRLAY 459
360 RHVRSDEP 367
460 --PRSGEE 465

RESULT 3

J0E0261 N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human

C:Species: Homo sapiens (man)
C:Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
C:Accession: J0E0261

R:Uchimura, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoka, T.

J: Biochem. 124, 670-678, 1998

A:Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of

A:Reference number: J0E0261; M01D:98391845; PMID:9722682

A:Accession: J0E0261

A:Molecule type: mRNA

A:Residues: 1-484 <UCH>

A:Cross-references: DDBJ:AB014679

C:Comment: This protein catalyzes the transfer of sulfur of sulfate from 3'-phosphoadenosine C:Superfamily: chondroitin 6-sulfotransferase

C:Keywords: sulfotransferase

Query Match 28.9%; Score 588.5; DB 2; Length 484;
Best Local Similarity 36.0%; Pred. No. 1,46-42;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;

[illegible]

Db 389 DSNQIYSTOKNSSQEFKWRPSIFPKLAQVVDACEPAMRLFGYKLASSAQLNTRISL 448
 Oy 376 L 376
 Db 449 L 449

RESULT 5

E95934

probable enzyme, C-terminal domain similar to sulfotransferase protein [Imported] - Sinc

C:Species: Sinorhizobium meliloti

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: E95934

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhrmester, J.; Chain, P.; Vorholter, F.J.; Hernan

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: E95934

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-307 <KOR>

A:Cross-references: GB:AL591985; PIDN:CAQ49141.1; PID:915140626; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSymb

R:Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.;

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure,

heubalt, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: SMD21237

A:Genome: plasmid

Query Match 8.2%; Score 168; DB 2; Length 307;

Best Local Similarity 22.7%; Pred. No. 9.2e-07;

Matches 83; Conservative 50; Mismatches 131; Indels 102; Gaps 16;

Oy 36 MKAPEREMHVLVLSMSGSSFEVQGLFGQHPDVF--YLMPEAMHWMTFKOSTAMLMHM 92
 Db 1 MPSQPVRL-AVIAGYGRSGSTILDLALGQHAAYVAGAGITSLTRHVRHNHCYCA----- 53
 Oy 93 AVRRLIRAVFLCDMSVDAVMEPRPRROSSLFQWENSALCSAPACDIIIPDEIIPRAHC 152
 Db 54 -----CGNAIRDCSFWSSVRR-----EMSDG-----ODPGLMEYXC 84
 Oy 153 RLCSQOPFEVEYKACR-----SYSHVLEKVEFFN-LQS----- 186
 b 85 AL---QKFEGLSMNTRLTSLGSGMLGKQFSLYIILTKRFLFSAMQSCSGQVYVDSKLPGR 141
 Oy 187 LYLPLKDPSLNLHVLVLRDPAVFRS-----RERTKGLMIDSRIVMGQHEOKLKED 240
 Db 142 AMAVAQIPGIDMRHILVDRGVAMSLKGYERDAKSG-----LQKEI 185
 Oy 241 OPYVMOVIOGSOLEIKTQSLPKAL-QERYLLVRYEDLARAVAOQTSRYEVEVGLLEFL 299
 Db 186 KPKSVFKALRWMS-VNLAVELYSRKLSGEVVMRREDESPDAVVMQIGTFLDLD-L 243
 Oy 300 PHLQTVNHNITRGKMG-----DHAFTNARDALNVSQAWMSLPEYKVSRLQKACGDA 353
 Db 244 SQVETSLEN---GEAMGQGVAGNRLRMNNSIALNDEITWRTMPPAQQVSPQRLGGM 300
 Oy 354 MNLLGY 359
 Db 301 LRRYGY 306

RESULT 6

G70729

hypothetical protein RV2267c - Mycobacterium tuberculosis (strain H37RV)

C:Species: Mycobacterium tuberculosis
 C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: G70729

R:Coile, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon

; Connor, R.; Davies, R.; Devlin, K.; Felwell, T.; Gentles, S.; Hamlin, N.; Holroyd,

Rajandream, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skellon, S.; Squares, S.

Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrett, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete geno

A:Reference number: A70500; MUID:98295987; PMID:9634230

A:Accession: G70729

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-388 <COD>

A:Cross-references: GB:Z77163; GB:AL123456; NID:93261610; PIDN:CAB00968.1; PID:e25507

A:Experimental source: strain H37RV

C:Genetics:

A:Gene: RV2267c

Query Match 5.7%; Score 116; DB 2; Length 388;

Best Local Similarity 19.3%; Pred. No. 0.036;

Matches 70; Conservative 67; Mismatches 111; Indels 114; Gaps 20;

Oy 45 VLVLSSWRSGSSFEVQGLF--GQH--PDVFLMEPAHWMTFKOSTAMLMHAYRDLIR 99
 Db 83 IFIVGHWRTGTLLHLLVVDRLHGTGPGYECIAP--HHFLT---TEM----- 125
 Oy 100 AVFLCDMSVDAVMEPRPRROSSLFQWENSALCSAPACDIIIPD-DEII-----PRAHC 152
 Db 126 -----FAPYVE-----FLSKHAKAMNMDLSLHNPDEDFEVMQGLPSPYL 167
 Oy 153 RLCSQOP-----FEVEKACRSYSHVLEKVEFFNLSY-----PLKDPSLNHI 200
 Db 168 TIAPNRPQYEEYIDLEQVAPREL--IMKRTLFEPQOYVFRKRTVILKNPHSRIRI 225
 Oy 201 -----VHLVRDPAVFRSRETKDLMIDSRIVMGQHEOKLKEDQRY--VM 246
 Db 226 KVLLEVPQAKFHHVDRPYVY-----PSYIHAKALYRIHGLOQPFEDGLD 273
 Oy 247 QVTCOSQLEIKTQSLPKALQ-EKYLVRVEDLARAVAOQTSRYEVEVGLF--FLPH 301
 Db 274 DKVYSTYVDLTKRDEGRELVDPRFELRLDLDGDEGLRRLYQRLGDEECYLP 333
 Oy 302 LOTWVHNITRGKMGDHA-FHTNARDALNVSQAWMSLPEYKVSRLQKACGDAMNLLGY- 359
 Db 334 LRQY-----LADHADYKTN-----YQLTVEQRAIVDEHMEIIDRYGYD 373
 Oy 360 RH 361
 Db 374 RH 375

RESULT 7

I40485

surfactin synthetase component I - Bacillus subtilis

N:Alternate names: competence protein srfAA; surfactin production protein srfAA; surf

C:Species: Bacillus subtilis

C:Date: 12-Aug-1996 #sequence_revision 12-Aug-1996 #text_change 03-Nov-2000

C:Accession: I40485; B69718; S35517; A37323; S46967; A43705; S34985

R:Cosmina, P.; Rodriguez, F.; de Ferra, F.; Grandi, G.; Perego, M.; Venema, G.; van S

Mol. Microbiol. 8, 821-831, 1993

A:Title: Sequence and analysis of the genetic locus responsible for surfactin synthe

A:Reference number: I40485; MUID:93360813; PMID:8355609

A:Accession: I40485

A:Status: translated from GB/EMBL/DBJ

A:Molecule type: DNA

A:Residues: 1-358 <RES>

A:Cross-references: EMBL:X70356; NID:9396480; PIDN:CAA49816.1; PID:9396481

A:Experimental source: strain w168 derivative of JH642

R:Kunst, F.; Ogasawara, N.; Moser, I.; Albertini, A.M.; Alloni, G.; Azavedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capiano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Errington, J.; Fabret, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Filtz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Gallizzi, A.; Galled
 lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, M.F.
 Koester, P.; Koningsstein, G.; Krogsh, S.; Kumano, M.; Kurita, K.; Lapius, A.; Lardinois,
 A:Authors: Lauber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Maueel
 Y, M.; Ogawa, K.; Oliveira, A.; Oudea, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portetelid
 Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanlon,
 A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Serot
 akuchl, M.; Tamakoshi, A.; Tanaka, T.; Terpstra, P.; Tognoni, A.; Tosato, V.; Uchiyama,
 T.; Winters, P.; Wilpat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K.
 A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.
 A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
 A:Reference number: A69580; MUID:98044033; PMID:984377
 A:Accession: B69718
 A:Status: nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-3588 <K0N>
 A:Cross-references: GB:299105; GB:AL009126; NID:g2632457; PIDN:CAB12142.1; PID:g2632634
 A:Experimental source: strain 168
 A:Title: Nucleotide sequence of 5' portion of *srfa* that contains the region required for
 Nucleic Acids Res. 21, 93-97, 1993
 A:Reference number: S35517; MUID:93181186; PMID:8441623
 A:Accession: S35517
 A:Status: nucleic acid sequence not shown; significant sequence differences
 A:Molecule type: DNA
 A:Cross-references: EMBL:D13262; NID:g216345; PID:g216346
 A:Experimental source: strain 168 trpC2
 A:Note: protein sequence not complete, the nucleotide sequence was submitted to the EMBL
 R:Nakano, M.M.; Magnusson, R.; Myers, A.; Curry, J.; Grossman, A.D.; Zuber, P.
 J. Bacteriol. 173, 1770-1778, 1991
 A:Title: *srfa* is an operon required for surfactin production, competence development, an
 A:Reference number: A37323; MUID:91154134; PMID:1847909
 A:Accession: A37323
 A:Status: nucleic acid sequence not shown; not compared with conceptual translation
 A:Molecule type: DNA
 A:Residues: 1-46, 'EV', '49-145', 'I', '147-150', 'L', '152', 'AN', '155-280', 'T', '282-307', 'SF', '310-384'
 A:Cross-references: GB:M59939; NID:g14366
 R:Nakano, M.M.; Xia, L.; Zuber, P.
 J. Bacteriol. 173, 5487-5493, 1991
 A:Title: Transcription initiation region of the *srfa* operon, which is controlled by the
 A:Reference number: A43705; MUID:91358326; PMID:1715856
 A:Contents: annotation
 R:Fabret, C.; Quentin, Y.; Gutseppl, A.; Busuttil, J.; Halech, J.; Denizot, F.
 submitted to the EMBL Data Library, March 1993
 A:Reference number: S46967
 A:Accession: S46967
 A:Molecule type: DNA
 A:Residues: 3249-3271, 'A', '3273-3316', 'R', '3318-3451', 'Y', '3453-3483', 'DE', '3486-3487', 'DAGU', '34
 A:Cross-references: EMBL:X72672; NID:g516358; PIDN:CAA51222.1; PID:g516359
 A:Experimental source: strain 168 trpC2
 C:Comment: This enzyme is one of several in the multienzyme complex that synthesizes the
 nee.
 C:Genetics:
 A:Gene: *srfa*
 A:Note: *srfa* is the first gene of the *srfa* operon
 C:Superfamily: surfactin synthetase; acetate-CoA ligase homology; acyl carrier protein h
 C:Keywords: antibiotic biosynthesis; carrier protein; phosphopantetheine; phosphoprotein
 F:507-935/Domain: acetate-CoA ligase homology <ACLI1>
 F:974-1042/Domain: acyl carrier protein homology <ACPI1>
 F:1043-1488/Domain: repeat <RPT1>
 F:1549-1993/Domain: acetate-CoA ligase homology <ACLI2>
 F:2011-2079/Domain: acyl carrier protein homology <ACPI2>
 F:2080-2327/Domain: repeat <RPT2>
 F:2589-3025/Domain: acetate-CoA ligase homology <ACLI3>
 F:3042-3109/Domain: acyl carrier protein homology <ACPI3>
 F:1006, 2043, 3074/Blinding site: phosphopantetheine (Ser) (covalent) #status predicted

Query Match 4.7%; Score 95; DB 2; Length 3588;
 Best Local Similarity 19.7%; Pred. No. 45;
 Matches 63; Conservative 48; Mismatches 111; Indels 98; Gaps 13;
 14 VSOMALFLFFHMTNHSLSMKAPPEMHVLLVSSMRSSGFVGLFGCHPDVFLME 73

Db 990 VERAQIFDNFEETGSHLAKMTLLTKIRK-----ETGIEIPQOFLFHEPTTALAE 1040
 74 PAMVMTTFKOSTAMMLHMAVRDLIRAVFLCDMSVDAYMEPPGRSSILFOWENSALC 133
 1041 EADH-----RESNAF-----AYIEPRKEHNPILHSHSEHS 1073
 134 SAPA-----CDIIPDEII-----PRAHRCILCS 157
 1074 SASSRRRESAIIHASSNSSECFRYSKAGARISGINPPIVIENTSFVLENSTPROKIH-VCV 1132
 158 QQPFVEVERACRSYSHVULKEVFENLQSLYPLKDPSP----NLHIV-----HLVRDP 208
 1133 DRIEMTERKGNDEDAIMASFVTFDLARA-PLFRIGLGLLEENRMLFLFDHHLTSDGV 1191
 209 AVFRSRERTKGLMID--SRIVGQHEOKLKKEDQPYVWQVTCOSQLEIYKTIQSIPRA 266
 1192 SI-----GIMBELARIVYGEQPLDLRQKYAVWQ--SRQAGEYKKDQYWK 1240
 267 L-QERYLYRYDRLARAV 284
 1241 VFAGELPVLQLLSDYPRPV 1260

RESULT 8
 T13288
 mel-41 protein - fruit fly (*Drosophila melanogaster*)
 C:Species: *Drosophila melanogaster*
 C:Date: 13-Aug-1999 #sequence_rev1sion 13-Aug-1999 #text_change 17-Nov-2000
 C:Accession: T13288
 R:Harl, K.L.; Santerre, A.; Sekelsky, J.J.; McKim, K.S.; Boyd, J.B.; Hawley, R.S.
 Cell 82, 815-821, 1995
 A:Title: The mel-41 gene of *D. melanogaster* is a structural and functional homolog of
 A:Reference number: Z11072; MUID:95401271; PMID:7671309
 A:Accession: T13288
 A:Status: preliminary; translated from GB/EMBL/DBJ
 A:Molecule type: DNA
 A:Residues: 1-2354 <HAP>
 A:Cross-references: EMBL:U34925; NID:g998351; PID:g998353; PIDN:AAC46881.1
 C:Genetics:
 A:Gene: mel-41
 A:Cross-references: FlyBase:Fgn0004367
 A:Insertions: 650/3; 748/3; 2313/3
 C:Function:
 A:Description: Involved in cell cycle checkpoint and meiotic recombination

Query Match 4.5%; Score 92.5; DB 2; Length 2354;
 Best Local Similarity 20.5%; Pred. No. 42;
 Matches 85; Conservative 52; Mismatches 147; Indels 131; Gaps 20;
 28 SHNRS---LSKKAQPERMHVLLVSSMRSSGFVGLFGCHPDVFLMEPAMHWMTFRK 84
 1058 NYNFASPOHPLNLTISDDFAVLATLSLCKGYQF--QONTKHYDSESL----- 1102
 85 STAMMLHMAVRDLIRAVFLCDMSVFDAYMEPPGRSSILFOWENSALCSAPACDIIPD 144
 1103 -----AIQETLA---TCGIS-----PKKKKQVLQW-----SLPA---RMQ 1133
 145 EIIPRAHRCILCSQPFVEVERAC--RSYSHVULKEVRF-----NLQS 186
 1134 LMEPMHSCYTCVHRSTGLQPLRGSHTSHNYTEMALMSRLDILPSSGKRILLSS 1193
 187 LYPPLKD-----PSLNLHIV-----HLVRDPAVFRSRERTKGLMIDSRIV 228
 1194 YKPCIKRDSNMLSTFPYLLHALLECTEQRNHIQDEPMVILQANEES-----SSSV 1246
 229 MCHQHEKLLKEDQPYVWQVTCOSQLEIYKTIQSIPKALQERYLYRYDRLARAVAYOTS 288
 1247 RGOEIGAIKEN-----AFKOFESRKYVAGI-KPLASTLVSDRKEDSSRVP----- 1291
 289 RMYEFGVLEFLPHLOTW-----VHNITGKMGSHAFPTNR---DALNVQAMRWSLP 339
 1292 RLAGKICAEILLDFLQRLWLRMQRIRHRSYG--CKPPTIDSNYRKTHIEFLN-----IP 1343

R.Bult, C.J.; White, O.; Olsen, G.J.; Zhou, L.; Fleischmann, R.D.; Sutton, G.G.; Blake, R.; Reich, C.T.; Overbeek, R.; Kirsnes, E.F.; Meisner, K.G.; Merrick, J.M.; Glodex, A.; Roon, J.D.; Sadow, P.W.; Hanna, M.C.; Cotton, M.D.; Roberts, K.M.; Hurst, M.A. Science 273, 1058-1073, 1996

A:Authors: Kaine, B.P.; Borodovsky, M.; Klenk, H.P.; Fraser, C.M.; Smith, H.O.; Moese, C.A.; Title: Complete genome sequence of the methanogenic archaeon, Methanococcus jannaschii A:Reference number: A63300; MUID:96337999; PMID:8688087

A:Accession: E64477

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-1847 <BUL>

A:Cross-references: GB:U67583; GB:L77117; NID:g2826409; PIDN:AB99433.1; PID:g1592072; C:Genetics:

A:Map position: REV1393176-1387633

A:Start codon: TTG

Query Match 4.5%; Score 91; DB 2; Length 1847;
Best Local Similarity 19.4%; Pred. No. 41;
Matches 46; Conservative 50; Mismatches 85; Indels 56; Gaps 9;

QY 93 AVDLIRAVFLCDMSYFADYMEGRPRQSSLPQWENSALCSAPACDIIPDEIIPRAC 152
D 269 ALRAFLRAYFDGDI-----EKHSIVLSTASKEMADLVVALRF 309
QY 153 RLCSQGPPEVEYKACRSYSHVL-----KEVRF-----NLQSLPLKDPPLNH 199
D 310 GIILAKLE-KVKNKNNKYHYHIVSNMIRFELNDIGFSQERKLLKLEIKDENPNND 368
QY 200 IYHLVDPRAVFRSRERTKGLMIDSRIVMGHEKLEKEDOPYVVOYICOSQL-EY- 257
D 369 VITI--DKKIRIYRDLKVKILTRD-----IEKDNMSYKCRKITQELLEYIY 415
QY 258 --KTQSLPLKQERYLL-----VXYEDLAPAVAGTSRMVEFVGLFPLQTVW 306
D 416 RLEELKEIKALEENLILMDVEAVERKEIAEKTGRSDRILEYIGRKRRPSIKNYI 472

RESULT 13

E86319
Probable flavonol sulfotransferase [imported] - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Mar-2001
C:Accession: E86319
R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alonso, Chlu, C.W.; Chung, M.K.; Com, L.; Conway, A.B.; Conway, A.R.; Creasy, T.H.; Dewar, K.; Jensen, N.F.; Hughes, B.; Huizart, L.
A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim, C.A.; Li, J.H.; Li, Y.; Liu, X.; Liu, S.X.; Liu, Z.A.; Luros, J.S.; Maitl, R.; Marzilli, Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallon, Ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venier, J.C.; Davis, R.W.
A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
A:Reference number: A86141; MUID:21016719; PMID:11130712
A:Accession: E86319
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-346 <STO>
A:Cross-references: GB:AE005172; NID:g9795597; PIDN:AAF98435.1; GSPDB:GN00141
C:Genetics:

Query Match 4.4%; Score 89.5; DB 2; Length 346;
Best Local Similarity 19.0%; Pred. No. 5.9;
Matches 72; Conservative 50; Mismatches 132; Indels 125; Gaps 17;

QY 28 SHNISLS---MKQPEMHVLYL---SSMRSSGFFGQLGQHPDVFYLMERAMHW 79
D 15 NHEILASSPSEFEKQKHQETIATLPKRGWRPPDPFE--YGGIMWLQPLLEGILHAQ 72
QY 80 MTF-----KOSTAMMLHNAVRLIRAVFLCDMSYFADYMEGRPRQSSLPQW 126

D 73 KFKAPRNDFFVCSYKRTGTTLKALTEA-----IANSKDVSTNPLKRNPREF-- 123
QY 127 ENSRALCSAPACDI---IPDEIIPRACRLCSQOPE---VVERACRSYSHVLK 177
D 124 -----VPIEIDPPFPSPVYLDNEGTLFSTHLPDLLESVVSGK----- 167
QY 178 EVAFENLQSLPLKDPSSLNLHVLVDRPRAVRS-----REKTKGLMIDSRIVMGO 231
D 168 -----IVYIWRPKDTFVSMTFAHKNRSQOQPVVSIIEAFDK 205
QY 232 HEOKLEKEDOPYVVOYICOSOLEIYKTIQSLPKALQERYLRYEDLARAPVAGTSRMV 291
D 206 YCGGLSAYG-PT-----LDHVLGYKAYQANP---DQLFLKYTEMRADLPYKRLA 254
QY 292 EFVGLF-----LPHQTVWNTNRGKGMDNAFHTNARDALNVSOA--- 333
D 255 EFVGYGFTKEEERGNVVEKVKLCSPET-LKNLEAKKGKDR---EDRPVAVANGAYFR 309
QY 334 -----WRMSLPYEKVSRL 346
D 310 KGRVGDWONYLTPPEVARI 328

RESULT 14

T07120
Probable cytochrome P450 CP7 - soybean
N:Contains: oxidoreductase (EC 1.-.-.-)
C:Species: Glycine max (soybean)
C:Date: 30-Apr-1999 #sequence_revision 30-Apr-1999 #text_change 21-Jul-2000
C:Accession: T07120
R:Schopfer, C.R.; Edel, J.
Mol. Gen. Genet. 258, 315-322, 1998
A:Title: Identification of elicitor-induced cytochrome P450s of soybean (Glycine max
A:Reference number: Z15931; MUID:98311068; PMID:9648734
A:Accession: T07120
A:Status: preliminary; translated from GB/EMBL/DBDUT
A:Molecule type: mRNA
A:Residues: 1-504 <SCH>
A:Cross-references: EMBL:Y10493; NID:g3334666; PIDN:CA71517.1; PID:g3334667
A:Experimental source: cultivar Harosoy 63, strain Merr
C:Superfamily: human cytochrome P450 CYP2D6; cytochrome P450 homology
C:Keywords: chromoprotein; electron transfer; heme; iron; metalloprotein; monooxygenase
F:403-466/Domain: cytochrome P450 homology <p45>
F:444/Binding site: heme iron (Cys) (axial ligand) #status predicted

Query Match 4.4%; Score 89.5; DB 2; Length 504;
Best Local Similarity 19.9%; Pred. No. 9.7;
Matches 70; Conservative 45; Mismatches 108; Indels 129; Gaps 16;

QY 68 VYIEMPRAMHWMTFQSTAMMLHNAVRLIRAVFLCDMSYFADYMEGRPRQSSLPQW 127
D 14 VFLLAH-WLV-KTYKQKSSHL-----PPGPMR----- 39
QY 128 NSRALCSAPACDIIPDEIIPRACRLCSQOPEVEVERACRSY---HVLYKEVFNIL 184
D 40 -----LP---IIGNHQALASLPDQALQVLRYKRGIMLQGLGISTVLV 83
QY 185 GSLYPLKDPSSLNLHVLVDRPRAVRSRERTKGLMIDSRIVMGHEKLEKEDOPY 244
D 84 SS--PKMAMEMKTHDVHFEVGRPO-----LLAPQPMVYGATDIAFAYGDYWR 129
QY 245 VMOYICOSOLEIYKTIQSLPKALQE---RLLYRYEDLARAPVAGTSRMVEFGL----- 296
D 130 QIRKICTTELLSAKRQSFSHLRQDENKRLQSIHSSAGSPIDLSGKFLSLGTYVSRNA 189
QY 297 -----EFL-----PHQTVWNTNRGKGMDNAFHTNARDALNVSOA--- 324
D 190 EKGENDQDEFFSVKATITNGGFEVDQMFSSLPK-LHLTRQAKAYEHV-HQRADKTL 247
QY 325 RDLNVSOAMRMSLPYEKVSRLQKACGDAMNLIGYRHVRSQEDQRLNLLDLL 376
D 248 EDILRKHM-----EKRTVKEGNG-----SEAEQEDLVVLL 279

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:39:29 ; Search time 9.47545 Seconds

(Without alignments)
1915.719 Million cell updates/sec

Title: US-09-645-078-2

Perfect score: 2038
Sequence: 1 MLPPKMKLLFLVSQMAIL.....EQRLNLLDLSTWVPEQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters: 127863

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Database: SwissProt_41.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	500.5	24.6	1 C6ST_CHICK	Q92179 gallus galli
2	116	5.7	1 YM67_MYCTU	Q50695 mycobacteri
3	104.5	5.1	1 NOEE_RHISN	P55472 rhizobium s
4	95	4.7	1 SRF1_BACSU	P27206 bacillus su
5	92	4.5	1 TSPB_CAEEL	Q20351 caenorhabdi
6	91.5	4.5	1 MACF_HUMAN	Q96p23 homo sapien
7	91.5	4.5	1 MAC4_HUMAN	Q96p22 homo sapien
8	91	4.5	1 239F_HUMAN	Q15777 homo sapien
9	91	4.5	1 MX3_RAT	P18590 rattus norv
10	89.5	4.4	1 C7D8_SOYBN	Q81974 glycline max
11	89.5	4.4	1 PKSM_BACSU	P40872 bacillus su
12	89	4.4	1 FAST_FLACH	P52837 flaveria ch
13	89	4.4	1 MPPA_RAT	P20069 rattus norv
14	89	4.4	1 CAPP_STRPY	Q9A0U7 streptococ
15	89	4.4	1 RRPD_PYWR	P17965 potato viru
16	88	4.3	1 F3ST_FLABI	P52835 flaveria bl
17	88	4.3	1 KNTC_HUMAN	P50748 homo sapien
18	87.5	4.3	1 TPSA_CAEEL	Q77081 caenorhabdi
19	87	4.3	1 CHRD_BRARE	Q57472 brachydanio
20	86.5	4.2	1 MOTA_MOUSE	P16332 mus musculi
21	86.5	4.2	1 DPOF_THERH	P30312 thermus the
22	86	4.2	1 SCH9_YEAST	P11792 saccharomy
23	85.5	4.2	1 BAL2_ARATH	Q9A8T4 arabidopsis
24	85	4.2	1 GLCE_CAEEL	P46555 caenorhabdi
25	85	4.2	1 CAPP_STRPY	Q8K873 streptococ
26	85	4.2	1 YDY2_SCHPO	Q03172 mus musculi
27	85	4.2	1 ZEP1_MOUSE	Q03172 mus musculi
28	84.5	4.1	1 SYT_LACLA	Q03172 mus musculi
29	84.5	4.1	1 YB4_YEAST	P35194 saccharomy
30	83.5	4.1	1 MSBA_CANAL	Q9A0T5 candida alb
31	83.5	4.1	1 GLND_PASMU	Q9C8H1 pasteurella
32	83	4.1	1 NOGG_XENLA	P49011 xenopus lae
33	82.5	4.0	1 CTDB_HUMAN	Q9H579 homo sapien

34	82.5	4.0	2549	1 FRAP_HUMAN	P42345 homo sapien
35	82.5	4.0	2549	1 FRAP_MOUSE	Q9J1n9 mus musculi
36	82.5	4.0	2549	1 FRAP_RAT	P42346 rattus norv
37	82	4.0	357	1 BUK_THERH	Q8R832 thermosacch
38	82	4.0	2352	1 MOKC_SCHPO	Q9U014 schizosacch
39	81.5	4.0	309	1 F3ST_FLABI	P52838 flaveria bl
40	81.5	4.0	311	1 RPL_VSVCO	P52836 flaveria ch
41	81.5	4.0	2109	1 RPL_VSVCO	P16379 vesicular s
42	81	4.0	534	1 THAS_PIG	P47787 sus acrofa
43	81	4.0	659	1 MX2_RAT	P18589 rattus norv
44	80.5	3.9	223	1 NOGG_CHICK	Q9J525 gallus galli
45	80.5	3.9	417	1 IE63_HSVSA	P13199 herpesvirus

ALIGNMENTS

RESULT 1	ID	Query Length	STANDARD	PRT	458 AA
1 C6ST_CHICK	Q92179	24.6	2549	1	FRAP_HUMAN
AC	Q92179	24.6	2549	1	FRAP_MOUSE
DT	01-NOV-1997 (Rel. 35, Created)				
DT	01-NOV-1997 (Rel. 35, Last sequence update)				
DT	01-NOV-1997 (Rel. 35, Last annotation update)				
DE	Chondroitin 6-sulfotransferase (EC 2.8.2.17) (C6ST).				
OS	Gallus gallus (Chicken).				
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;				
OC	Archosauria; Aves; Neognathae; Galliformes; Phasianidae; Phasianinae;				
OC	Gallus.				
OX	NCBI_TaxID:9031;				
RN	[1]				
RP	SEQUENCE FROM N.A., AND SEQUENCE OF 38-44; 46-52 AND 186-192.				
RC	SPRAIN-White leghorn; TISSUE-Embryonic chondrocytes;				
RX	MEDLINE-95355490; PubMed-7629189;				
RA	Fukuta M., Uchimura K., Nakashima K., Kato M., Kinata K.,				
RA	Shinomura T., Habuchi O.;				
RT	"Molecular cloning and expression of chick chondrocyte chondroitin 6-				
RT	sulfotransferase.";				
RL	J. Biol. Chem. 270:18575-18580(1995).				
CC	-1- FUNCTION: CATALYZES THE SULFATION OF CHONDROITIN AND KERATAN				
CC	SULFATE.				
CC	-1- CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + chondroitin -				
CC	adenosine 3',5'-bisphosphate + chondroitin 6'-sulfate.				
CC	-1- SUBCELLULAR LOCATION: TYPE II MEMBRANE PROTEIN. GOLGI MEMBRANE				
CC	(BY SIMILARITY).				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL: D49915; BAA08655.1;				
DR	PIR: A57397; A57397				
DR	InterPro: IPR000863; Sulfotransferase.				
DR	Transferrase; Transmembrane; Signal-anchor; Glycoprotein; Golgi stack.				
KW	Transferrase; Transmembrane; Signal-anchor; Glycoprotein; Golgi stack.				
FT	DOMAIN 1 23				
FT	TRANSMEM 24 37				
FT	FT				
FT	DOMAIN 38 458				
FT	CARBOHYD 62 62				
FT	CARBOHYD 73 73				
FT	CARBOHYD 95 95				
FT	CARBOHYD 236 236				
FT	CARBOHYD 399 399				
FT	CARBOHYD 443 443				
FT	SEQUENCE 458 AA; 52253 MW; C9A3B7DDA5086F0C CRC64;				
SO	SEQUENCE				
Query Match	24.6%; Score 500.5; DB 1; Length 458;				
Best Local Similarity	33.2%; Pred. No. 4; 1e-35;				

Matches 120; Conservative 75; Mismatches 133; Indels 33; Gaps 13;

QY 34 LSMKAQPERHVLVLSWRSQSSFYVQGLFQHPDVEYLMPEAMHV--WMTFKOSTAMMLH 91
 DB 104 LGIAPRPRHVLMTATRTGSSFYGEFFNNOGNFIYLEPMLHIERVTTFEGGANNAG 163
 QY 92 NAV--RDLIRAVFLCDMSVFDAYMEGPPRQ--SSIFQWENSRALCSAPACDIIPQ-DEI 146
 DB 164 SALVRYRDVYQQLDLDVLTLEESFISPAPEHNTALFRGSSHSLSCEEPVC--TPSLKRV 221
 QY 147 IRRACR--LLCSQPFEEVYKACRSYSHVLKEVREFNQSILYPLKDSLNLHAYLR 205
 DB 222 FEKYCKNRCRCPPLMTITLAEACRKHAKLTVRLOLEFQPLAEDRLDRITQIVR 281
 QY 206 DRAVFRSREKRGDLMDISRLVWGQHECKL-----KEDPYVMQVYCS-QLE 255
 DB 282 DRAVLVSR-----MV---AFSGYSESKKMAEGEAPLQDE-VQLRGNCESIRLS 330
 QY 256 IKTTOSLPKALQERYLLVRYEDLARAPVQTSRMTEFVGLFELPLQTVWHNITRGKM 315
 DB 331 AELGRQ--PRWLRGRMLVRYEDVAPLRKALEMREFAGIHPTQVCEMIRANTQAP-Q 388
 QY 316 GDHAFHTNARDALNVSQAWMSLPEYKVSRLQKACGDANLLGVRHVRSEQRNLLDL 375
 DB 389 DSGNYSYQKNSSEGEFEKRFESIFRLQAVVODACEPARMLFGYKLASSQELTNSLSL 448
 QY 376 L 376
 DB 449 L 449

RESULT 2
 YH67_MYCTU STANDARD; PRT; 388 AA.
 AC Q50695;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein Rv2267c.
 GN Rv2267c OR M2329 OR MTC339.43.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinomycetales;
 OC Corynebacterineae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID-1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RA MEDLINE-98295987; PubMed-9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekela E.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy J.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sultun J.E., Taylor K., Whitehead S., Barrall B.G.,
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544(1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., DeBoy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.,
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains."
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 DR EMBL: Z77163; CAB00968.1;
 DR EMBL: AE007076; NAK46611.1;
 DR PIR: G70729; G70729.
 DR TIGR: MT2329;
 DR Tuberculist: Rv2267c;
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 388 AA; 46062 MW; 5DE0263275A9B24 CRC64;

Query Match 5.7%; Score 116; DB 1; Length 388;
 Best local similarity 19.3%; Pred. No. 0 018;
 Matches 70; Conservative 67; Mismatches 111; Indels 114; Gaps 20;

QY 45 VLVLSWRSQSSFYVQGLF--GQH--PDVEYLMPEAMHVMTFKOSTAMMLHNAVVDLIR 99
 DB 83 IFVGHMRTGTLTLLHLLVVDVDRHTGPTGYECLAP--HHFLT---TEM----- 125
 QY 100 AVFLCDMSVFDAYMEGPPRQSSLPQWMSRALCSAPACDIIPQ-DEIT-----PRAHC 152
 DB 126 -----FAPVE-----FLVSKHRADNNDSLHHPQDEDFWCMQGLSPYLL 167
 QY 153 RLCSQGP-----FEVYKACRSYSHVLKEVREFNLSLY-----PLKDPSSLNLI 200
 DB 168 TIAFPNRPPOYEXYEDLEQVAPRELE--IMKRTLFYVQVYFRKRTYILNPTISFRI 225
 QY 201 -----VHLVRDRAVFRSREKRGDLMDISRLVWGQHECKLKEDEPY--VN 246
 DB 226 KVLLEFPQAKFTHIYRDYVYV-----PSYIHLKALYRIHGLQPTPDGLD 273
 QY 247 QVYCSQSLTYKTQSLPKALQ--ERYLLVRYEDLARAPVQTSRMTEFVGLF-----FLPH 301
 DB 274 DKVYSTYVDLYRLDGRGLVDPTRELYETEDLIDDPGQRLRLQHLGLDGEYCLPR 333
 QY 302 LQTVWHNITRGKMGDHA-FHTNARDALNVSQAWMSLPEYKVSRLQKACGDANLLGVR 359
 DB 334 LRQY-----LADHADYKTN-----YQLVVEDQALYDEHMGELIDRYGYD 373
 QY 360 RH 361
 DB 374 RH 375

RESULT 3
 NOEF_RHISN STANDARD; PRT; 419 AA.
 AC P55472; P72326;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 15-DEC-1998 (Rel. 37, Last annotation update)
 DE Nodulation protein noe (EC 2.8.2.-).
 GN NOE OR Y4HB.
 OS Rhizobium sp. (strain NGR234).
 OS Plasmid sym pNGR234a.
 OC Bacteria; Proteobacteria; Alphaproteobacteria; Rhizobiales;
 OC Rhizobiaceae; Rhizobium/Agrobacterium group; Rhizobium.
 OX NCBI_TaxID-394;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-97361801; PubMed-9218762;
 RA Hanin M., Jabbour S., Quesada-Vincens S., Freiberg C., Perret X.,
 RA Broughton W.J., Fellay R.,
 RT "Sulphation of Rhizobium sp. NGR234 Nod factors is dependent on noe,
 RT a new host-specificity gene."
 RL Mol. Microbiol. 24:1119-1129(1997).
 RN [2]
 RP SEQUENCE FROM N.A.

RA MEDLINE-97305956; PubMed-9163424;
RA Freiberg C.A., Fellay R., Broughton W.J., Rosenthal A.,
RA Perret X.;
RT "Molecular basis of symbiosis between Rhizobium and legumes."
RL Nature 387:394-401(1997).
CC -1- FUNCTION: REQUIRED FOR THE FORMATION OF SULFATED NOD FACTOR.
CC PROPOSED TO TRANSFER ACTIVATED SULFATE (PAPS) TO THE Fucose
CC OF THE NOD FACTOR.
CC -1- SIMILARITY: LIMITED TO NODH AND TO C.ELEGANS F4269.8.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; Y09415; CAA70569.1; -
DR EMBL; AE000076; AAB91690.1; -
DR InterPro: IPR000863; Sulfoltransferase.
DR Pfam: PF00685; Sulfoltransfer 1.
KW Nucleation; Transferase; Plasmid.
SQ SEQUENCE 419 AA; 46569 MW; 848C48E0416AAALF CRC64;

Query Match 5.1%; Score 104.5; DB 1; Length 419;
Best Local Similarity 21.5%; Pred. No. 0.19;
Matches 78; Conservative 48; Mismatches 134; Indels 103; Gaps 18;

QY 34 LSKKADPERMNVILVSSWSSGSSFGQLFGQHPDVFYLMPEAMHVMTFKOSTAMLMHA 93
DB 1 MSRDVSPPLPICFLGIPRSGTLLAHLQHPDITAPRP-----WL----- 43
QY 94 VRLIRAVFLCDMSVPAYVPEPRRROSSIFOMENSALCSAPACDIIPODEIT----- 147
DB 44 -----MLAEAFGRVDHHPA-----GSLVQVACEPLGRIDRISVRVF 84
QY 148 -PRAHCLTCSQPFEEVERKACRSYSHVLEKEVFEMQSLYPLKPSINLHIVHVRD 206
DB 85 ADAAVSOYLAAAKRRTIDKTPRW--NVID-----YLHSLYF--EAP-----HL-LLRN 130
QY 207 PRAVFSRERTKGDMLDSRIYVNGHEQKLKKEDQPYVYVQVICOQO-----LE 255
DB 131 PYAIAASLTKSTWCVPEVSEKCP-----PTVSQIALELVTGTPTAAVALA 174
QY 256 IYTTISLPRALDER-----YLVRYEDLARAPYAQTSRM-----YEEVGLFELPHLTGW 306
DB 175 LADLVGLPALAMQGRHRTQVRYELVERPDEIQRLAAGLGYDPAGIVFAGVEQT-- 232
QY 307 HNTTCKGMGD-HAFHTNARDALNVSQAMRMSLPYEKYSRLQKACG--DANMLIGY-----R 360
DB 233 -EYLRLLSFFGRRLKKKAVDNRSV--ETWTELTIEEMQVITDVLGADLVLELGEYSLO 290
QY 361 HVR 363
DB 291 HAR 293

RESULT 4
SRFL_BACSU
ID SRFL_BACSU STANDARD; PRT; 3588 AA.
AC P27206;
DT 01-AUG-1992 (Rel. 23, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Surfactin synthetase subunit 1.
GN SRFAA OR SRFA1 OR SRFA.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;

RA MEDLINE-93181186; PubMed-8441623;
RA Funa S., Fujishima Y., Corbell N., D'Souza C., Nakano M.M.,
RA Zuber P., Yamane K.;
RT "Nucleotide sequence of 5' portion of srfA that contains the region
RT required for competence establishment in Bacillus subtilis."
RL Nucleic Acids Res. 21:93-97(1993).
RN [2]
RN SEQUENCE FROM N.A.
RC STRAIN=168 / JH642;
RX MEDLINE-93360813; PubMed-8355609;
RA Cosmina P., Rodriguez F., de Ferra F., Grandi G., Perego M.,
RA Venema G., van Sinderen D.;
RT "Sequence and analysis of the genetic locus responsible for surfactin
RT synthesis in Bacillus subtilis."
RL Mol. Microbiol. 8:821-831(1993).
RN [3]
RN SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE-97124189; PubMed-8969502;
RA Yamane K., Kumano M., Kurita K.;
RT "The 25 degrees-36 degrees region of the Bacillus subtilis
RT chromosome: determination of the sequence of a 146 kb segment and
RT identification of 113 genes."
RL Microbiology 142:3047-3056(1996).
RN [4]
RN SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE-98044033; PubMed-9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Bertolo M.G., Bessieres P., Bolotin A., Borchert S.,
RA Brouillet S., Boursier L., Brans A., Braun M., Brigelli S.C., Bron S.,
RA Chai S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denlioz F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Enlian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Hatched J., Harwood C.R., Henaut L.,
RA Hilbert H., Holsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klaert-Blanchard M., Klein C.,
RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Meliadi R.P., Mizuno M., Moesti D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
RA Patro V., Pohl T.M., Portetelle D., Portocarrillo S., Prescott A.M.,
RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadate Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Solido B.,
RA Socotkin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Yamakoshi A., Tanaka T., Terpiltra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
RT subtilis."
RL Nature 390:249-256(1997).
RN [5]
RN SEQUENCE OF 1-460 FROM N.A.
RX MEDLINE-91154134; PubMed-1847909;
RA Nakano M.M., Magnuson R., Myers A.M., Curry J., Grossman A.D.,
RA Zuber P.;
RT "srfA is an operon required for surfactin production, competence
RT development, and efficient sporulation in Bacillus subtilis."
RL J. Bacteriol. 173:1770-1778(1991).
RN [6]
RN SEQUENCE OF 1-64 FROM N.A.
RC STRAIN=168;
RX MEDLINE-95219080; PubMed-7704255;
RA Fujishima Y., Yamane K.;
RT "A 10 kb nucleotide sequence at the 5' flanking region (32 degrees)

RT of srfaA of the Bacillus subtilis chromosome. ;
 RL Microbiology 141:277-279(1995).
 RN [7]
 RP SEQUENCE OF 1-38 FROM N.A.
 RX MEDLINE-91358326; PubMed-1715856;
 RA Nakano M.M., Xia L., Zuber P.;
 RT "Transcription Initiation region of the srfa operon, which is
 controlled by the comp-coma signal transduction system in Bacillus
 subtilis. ;"
 RL J. Bacteriol. 173:5487-5493(1991).
 CC -1- FUNCTION: THIS PROTEIN IS A MULTIFUNCTIONAL ENZYME ABLE TO
 CC ACTIVATE AND POLYMERIZE THE AMINO ACIDS LEU, GLU, ASP AND VAL.
 CC ACTIVATION SITES FOR THESE AA CONSIST OF INDIVIDUAL DOMAINS.
 CC -1- CORRECTOR: CONTAINS 3 COVALENTLY BOUND PHOSPHOPANTHETINES.
 CC -1- PATHWAY: Cyclic peptide antibiotic surfactin biosynthesis.
 CC -1- SIMILARITY: BELONGS TO THE ATP-DEPENDENT AMP-BINDING ENZYME
 CC FAMILY.
 CC -1- SIMILARITY: Contains 3 acyl carrier domains.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 DR EMBL; D13262; BAA02522.1; -
 DR EMBL; X70356; CAA49816.1; -
 DR EMBL; D50453; BAA08982.1; -
 DR EMBL; Z99105; CAB12142.1; -
 DR EMBL; M59639; AAA22815.1; -
 DR EMBL; D30762; BAA21034.1; -
 DR EMBL; M64702; AAA22816.1; -
 DR PIR; I40485; I40485.
 DR HSSP; P14687; JAMU.
 DR Subtilist; BG10168; srfaA.
 DR InterPro; IPR000873; AMP-bind.
 DR InterPro; IPR001242; Condensatn.
 DR InterPro; IPR006163; PP_bind.
 DR InterPro; IPR006162; Pantne_attach.
 DR Pfam; PF00501; AMP-binding; 3.
 DR Pfam; PF00668; Condensation; 4.
 DR Pfam; PF00550; pp-binding; 3.
 DR PRINTS; PRO0154; AMPBINDING.
 DR PROSITE; PS00012; PHOSPHOPANTHETINE; 2.
 DR PROSITE; PS00455; AMP BINDING; 3.
 DR PROSITE; PS00075; ACP DOMAIN; 3.
 KW Ligase: Antibiotic biosynthesis; Phosphopantetheine; Sporulation;
 KW Multifunctional enzyme; Repeat; Complete proteome.
 FT REPEAT ? 1047
 FT REPEAT ? 2084
 FT DOMAIN ? 3115
 FT DOMAIN 976 1043
 FT DOMAIN 2013 2080
 FT DOMAIN 3044 3110
 FT BINDING. 1006 1006
 FT BINDING. 2043 2043
 FT BINDING. 3074 3074
 FT CONFLICT 146 146
 FT CONFLICT 151 151
 FT CONFLICT 165 165
 FT CONFLICT 281 281
 FT CONFLICT 460 460
 FT CONFLICT 540 540
 FT CONFLICT 562 562
 FT CONFLICT 639 640
 FT CONFLICT 644 644
 FT CONFLICT 649 649
 FT CONFLICT 1026 1026
 FT CONFLICT 1065 1114

FT CONFLICT 1130 1131
 FT CONFLICT 1162 1162
 FT CONFLICT 1456 1456
 FT CONFLICT 1848 1848
 FT CONFLICT 1892 1892
 FT CONFLICT 1909 1909
 FT CONFLICT 1971 1978
 FT CONFLICT 2050 2050
 FT CONFLICT 2216 2216
 FT CONFLICT 2263 2264
 FT CONFLICT 2289 2289
 FT CONFLICT 2347 2347
 FT CONFLICT 2426 2426
 FT CONFLICT 2609 2610
 FT CONFLICT 2684 2686
 FT CONFLICT 2756 2756
 FT CONFLICT 2897 2898
 FT CONFLICT 3026 3026
 FT CONFLICT 3097 3097
 FT CONFLICT 3272 3272
 FT CONFLICT 3317 3317
 FT CONFLICT 3452 3452
 FT CONFLICT 3484 3496
 FT CONFLICT 3500 3501
 FT CONFLICT 3508 3508
 SQ SEQUENCE 3588 AA; 402072 MW; BC02PB15D7F1FDB CRC64;
 Query Match 4.7%; Score 95; DB 1; Length 3588;
 Best Local Similarity 19.7%; Pred. No. 19; Indels 98; Gaps 13;
 Matches 63; Conservative 48; Mismatches 111;
 QY 14 VSOMATIALEFHYHYSNHSLSKAPDEMHVLYLSMSGGSFVQLFGHPDYVLYME 73
 DB 990 VERAGIFDNFEFFEGHSLKAMTLTKIKH-----ETGIEIPQGFLEHPTITLALME 1040
 QY 74 PAMHWMTFKOSTAMLMHNAVRLINAVFLCMSVDAYMEQPPROSSLFQWENSRALC 133
 DB 1041 EADH-----RESKAF-----AVIEPAEKQEHYPLHMSSEHS 1073
 QY 134 SAPA-----CDIIPODEII-----PRAHCLICS 157
 DB 1074 SASRRRESAIIHSSNSGCFRYSKAGALISGINTPPRIETFSYLENSTPROKIH-VCV 1132
 QY 158 QQFEVEYERACRSYSHVLEKVEFNMQLSLYPLNDPSL-----NLHIV-----HLVRDR 208
 DB 1133 DFNIEIMERGSDAIDMASFVRTFDLANA-PLFRIGLGLDENRMLLFDMHHLISDGV 1191
 QY 209 AVFRSEKRTKGLMDI--SRIVMGHEQKIKKEDQPYIYQVYCQSLITYKTIOSLPRA 266
 DB 1192 SI-----GIMLEELARIYKGEQLPDLRLQYKDAYAWQ--SRQAAEGYKRDAYWKE 1240
 QY 267 L--QERYLLVREYEDLARAVP 284
 DB 1241 VRAGELPVLQQLSDYPRPV 1260
 RESULT 5
 TPSS CAEEL STANDARD; PRT; 359 AA.
 AC Q20351;
 DT 30-MAY-2000 (Rel. 39, Created)
 DT 30-MAY-2000 (Rel. 39, Last sequence update)
 DT 30-MAY-2000 (Rel. 39, Last annotation update)
 DE Putative protein-tyrosine sulfotransferase (EC 2.8.2.20)
 DE (Tyrosylprotein sulfotransferase) (TPST).
 GN P4269.8.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidae;
 OC Rhabditidae; Peloderinae; Caenorhabditis.
 OX NCBI_TaxID=6239;
 RN [1]
 SEQUENCE FROM N.A.

QY 337 SLPEKVSRLQKACGDAMNLLGYRHRVSEOEORILL 373
 11 11 11 11 11 11 11 11 11 11
 Db 2032 SLA-ERSSLQKALAGS-----QSVESLESL 2058

RESULT 7
 MAC4_HUMAN STANDARD; PRT; 5938 AA.
 ID MAC4_HUMAN
 AC 096PK2; 08WXY1;
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last sequence update)
 DE Microtubule-actin crosslinking factor 1, isoform 4.
 GN MACF1 OR ACFT OR ABP620 OR KIAA0465 OR KIA1251.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., ALTERNATIVE SPLICING, AND TISSUE SPECIFICITY.
 RX MEDLINE-21833812; PubMed-11845288;
 RA Gong T.-W.L., Bestiri C.G., Lomax M.I.;
 FT "MACF1 gene structure: a hybrid of plectin and dystrophin.";
 RL Mamm. Genome 12:852-861(2001).
 CC -1- FUNCTION: May play a role in cross-linking cytoskeletal proteins
 by binding intermediate filaments to the N-terminal plectin
 repeats and microtubules to the C-terminal.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
 CC -1- ALTERNATIVE PRODUCTS:
 Event-Alternative splicing; Named isoforms=4;
 CC Name=4;
 CC IsoId=096PK2-1; Sequence-Displayed;
 CC Name=1;
 CC IsoId=09UPN3-1; Sequence-External;
 CC Name=2;
 CC IsoId=09UPN3-2; Sequence-External;
 CC Name=3;
 CC IsoId=09UPN3-3; Sequence-External;
 CC -1- TISSUE SPECIFICITY: Expressed in heart, lung, pituitary and
 placenta, not found in brain, kidney, liver, pancreas or skeletal
 muscle.
 CC -1- SIMILARITY: Belongs to the plectin or cytolinker family.
 CC -1- SIMILARITY: Contains 2 EF-hand calcium-binding domains.
 CC -1- SIMILARITY: Contains 19 plectin repeats.
 CC -1- SIMILARITY: Contains 32 spectrin repeats.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC or send an email to license@isb-sib.ch.
 DR EMBL: AF317696; AAL09459.1; -;
 DR EMBL: AF325341; AAL38997.1; -;
 DR EMBL: AF325333; AAL38997.1; JOINED.
 DR EMBL: AF325334; AAL38997.1; JOINED.
 DR EMBL: AF325335; AAL38997.1; JOINED.
 DR EMBL: AF325336; AAL38997.1; JOINED.
 DR EMBL: AF325339; AAL38997.1; JOINED.
 DR EMBL: AF325340; AAL38997.1; JOINED.
 DR GO: GO:0008017; F-microtubule binding; ISS.
 DR InterPro: IPR002048; EF-hand.
 DR InterPro: IPR003108; GAS2.
 DR InterPro: IPR001101; Plectin_repeat.
 DR InterPro: IPR002017; Spectrin.
 DR Pfam: PF00036; ehand; 2.
 DR Pfam: PF02187; GAS2; 1.
 DR Pfam: PF00681; Plectin; 11.
 DR Pfam: PF00435; Spectrin; 26.
 DR ProDom: PD000012; EF-hand; 1.
 DR SMART: SM00054; Efn; 2.

DR SMART: SM00243; GAS2; 1.
 DR SMART: SM00250; PLEC; 19.
 DR SMART: SM00150; SPEC; 32.
 DR PROSITE: PS00018; EF_HAND; 2.
 DR Cytochrome: Calcium; Calcium-binding; Repeat; Coiled coil;
 KW Alternative splicing.
 FT DOMAIN 1830 1936 COILED COIL (POTENTIAL).
 FT DOMAIN 2001 2192 COILED COIL (POTENTIAL).
 FT DOMAIN 2282 2345 COILED COIL (POTENTIAL).
 FT DOMAIN 2477 2507 COILED COIL (POTENTIAL).
 FT DOMAIN 2541 2654 COILED COIL (POTENTIAL).
 FT DOMAIN 2686 2814 COILED COIL (POTENTIAL).
 FT DOMAIN 2887 2919 COILED COIL (POTENTIAL).
 FT DOMAIN 3046 3197 COILED COIL (POTENTIAL).
 FT DOMAIN 3262 3503 COILED COIL (POTENTIAL).
 FT DOMAIN 3632 3666 COILED COIL (POTENTIAL).
 FT DOMAIN 3746 3779 COILED COIL (POTENTIAL).
 FT DOMAIN 3920 3984 COILED COIL (POTENTIAL).
 FT DOMAIN 4098 4168 COILED COIL (POTENTIAL).
 FT DOMAIN 4288 4308 COILED COIL (POTENTIAL).
 FT DOMAIN 4354 4386 COILED COIL (POTENTIAL).
 FT DOMAIN 4397 4433 COILED COIL (POTENTIAL).
 FT DOMAIN 4469 4489 COILED COIL (POTENTIAL).
 FT DOMAIN 4586 4720 COILED COIL (POTENTIAL).
 FT DOMAIN 4845 4880 COILED COIL (POTENTIAL).
 FT DOMAIN 4910 4939 COILED COIL (POTENTIAL).
 FT DOMAIN 4970 5000 COILED COIL (POTENTIAL).
 FT DOMAIN 5409 5437 COILED COIL (POTENTIAL).
 FT DOMAIN 5546 5569 COILED COIL (POTENTIAL).
 FT CA_BIND 5598 5610 EF_HAND 1 (POTENTIAL).
 FT CA_BIND 5634 5646 EF_HAND 2 (POTENTIAL).
 FT REPEAT 12 49 PLECTIN 1.
 FT REPEAT 53 88 PLECTIN 2.
 FT REPEAT 89 126 PLECTIN 3.
 FT REPEAT 130 164 PLECTIN 4.
 FT REPEAT 166 202 PLECTIN 5.
 FT REPEAT 203 240 PLECTIN 6.
 FT REPEAT 243 278 PLECTIN 7.
 FT REPEAT 279 316 PLECTIN 8.
 FT REPEAT 318 354 PLECTIN 9.
 FT REPEAT 725 762 PLECTIN 10.
 FT REPEAT 763 800 PLECTIN 11.
 FT REPEAT 801 838 PLECTIN 12.
 FT REPEAT 839 876 PLECTIN 13.
 FT REPEAT 897 934 PLECTIN 14.
 FT REPEAT 935 972 PLECTIN 15.
 FT REPEAT 975 1011 PLECTIN 16.
 FT REPEAT 1012 1049 PLECTIN 17.
 FT REPEAT 1121 1158 PLECTIN 18.
 FT REPEAT 1159 1196 PLECTIN 19.
 FT REPEAT 1925 2032 PLECTIN 1.
 FT REPEAT 2052 2160 PLECTIN 2.
 FT REPEAT 2211 2313 PLECTIN 3.
 FT REPEAT 2320 2430 PLECTIN 4.
 FT REPEAT 2437 2543 PLECTIN 5.
 FT REPEAT 2550 2652 PLECTIN 6.
 FT REPEAT 2659 2758 PLECTIN 7.
 FT REPEAT 2765 2896 PLECTIN 8.
 FT REPEAT 2903 3008 PLECTIN 9.
 FT REPEAT 3015 3119 PLECTIN 10.
 FT REPEAT 3126 3229 PLECTIN 11.
 FT REPEAT 3236 3339 PLECTIN 12.
 FT REPEAT 3346 3446 PLECTIN 13.
 FT REPEAT 3453 3555 PLECTIN 14.
 FT REPEAT 3562 3664 PLECTIN 15.
 FT REPEAT 3671 3775 PLECTIN 16.
 FT REPEAT 3782 3884 PLECTIN 17.
 FT REPEAT 3891 3993 PLECTIN 18.
 FT REPEAT 4000 4102 PLECTIN 19.
 FT REPEAT 4109 4211 PLECTIN 20.
 FT REPEAT 4218 4320 PLECTIN 21.
 FT REPEAT 4327 4428 PLECTIN 22.
 FT REPEAT 4438 4544 PLECTIN 23.

```

FT REPEAT 4551 4653 SPECTRIN 24.
FT REPEAT 4660 4763 SPECTRIN 25.
FT REPEAT 4770 4872 SPECTRIN 26.
FT REPEAT 4879 4982 SPECTRIN 27.
FT REPEAT 4989 5091 SPECTRIN 28.
FT REPEAT 5098 5201 SPECTRIN 29.
FT REPEAT 5208 5309 SPECTRIN 30.
FT REPEAT 5316 5418 SPECTRIN 31.
FT REPEAT 5425 5555 SPECTRIN 32.
FT CONFLICT 1712 1712 P -> S (IN REF. 1: AAL38997).
SQ SEQUENCE 5938 AA; 670134 MW; B8784112752DA004 CRC64;

Query Match 4.5%; Score 91.5; DB 1; Length 5938;
Best Local Similarity 22.2%; Pred. No. 74;
Matches 88; Conservative 66; Mismatches 128; Indels 115; Gaps 24;

QY 2 LIPKMKLLFLVSGMALLALFEHMYSHNSSLKAKQPERHNVLYLSSMRGSSFYGL 61
DB 2254 LLSQONFILTATQSKA---FLDQGHNL-----PEEQML-----QOKLGL 2294
QY 62 FGOHDPVYLLMPAHVMTFKOSTAMMLHM-AVBDLIRAVFLCDMSYFADAYMEPGPRQ 120
DB 2295 KEQYS-----TSLAQSEALFKQVOTLQDELQK-FLQHKKEFWLE---RSE 2337
QY 121 SLPFGWNSRALCSAPACDIIPQDEIIPRAHRLCSQPF--EYVERKCRSYSH----- 173
DB 2338 KEL---ENMHKGGSSP-----FTLP-----SLKRGGSFSEVDY-----SHKGLR 2375
QY 174 -VLEKEVFNFLOSILYPLKPSLNLHVLVDRPRAVRSRERTKGLMIDSR----- 226
DB 2376 FTISGQAVLDMENFKSGKERS---EIGNLVKD-----KLMDERTYALMSKCTRGS 2427
QY 227 ---YMGQHEQKLKEDQPYVMQYICQSLQ-ITYKTQSIPKALQERYLLVREEDLAR 281
DB 2428 HLMMLLGGYHOPQNSDAQMMA-CENAVKELSLDVTASDPVQLQED----- 2475
QY 282 APVACTSRKRYEGVLEFLP--HLOTWVNI--TRCKMGMDHFNINARA-LNVQAWNR 336
DB 2476 --LATTKQIQEELAHQVVEKIQVARDIMEIEGPARDHNVQETTDLSLHQSLSY 2533
QY 337 SLPEKVSRLQKACGDANMLLGYRHVRSQEOQRNL 373
DB 2534 SLA-ERSSILQKAIQS-----QSVQESLESLL 2560

RESULT 8
339F_HUMAN
ID 239F_HUMAN STANDARD; PRT; 294 AA.
AC 015777;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-SEP-2003 (Rel. 42, Last annotation update)
DE Fetal brain protein 239 (239FB).
GN C11ORF8.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=95080775; PubMed=7527372;
RA Schwartz F., Neve R., Eisenman R., Gessler M., Bruns G.;
RT "A WAGR region gene between PAX-6 and FSHB expressed in fetal brain.";
RL Hum. Genet. 94:658-664(1994).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=96115606; PubMed=8666403;
RA Schwartz F., Eisenman R., Knoll J., Gessler M., Bruns G.;
RT "cDNA sequence, genomic organization, and evolutionary conservation
of a novel gene from the WAGR region.";
RL Genomics 29:526-532(1995).
RN [3]
RP SEQUENCE FROM N.A.

```

```

RC TISSUE=Colon, Kidney, and Stomach;
RX MEDLINE=22388257; PubMed=12477932;
RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA Klausner R.D., Collins F.S., Wagner L., Shemen C.M., Schuler G.D.,
RA Altschul S.F., Zeeberg B., Buetow K.H., Schefer C.F., Bhat N.K.,
RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA Brownstein M.J., Ueda T.B., Toshiyuki S., Carninci P., Bouffard G.G.,
RA Raha S.S., Loguercio N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA Bosak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
RA Villalón D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA Fahy J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA Butlerfield I.S.N., Krzywinski M.I., Skalska U., Smallos D.E.,
RA Scherch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT "Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences.";
RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC -1- TISSUE SPECIFICITY: EXPRESSED PREDOMINANTLY IN FETAL BRAIN.
CC -1- SIMILARITY: BELONGS TO THE UPF0046 FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC at the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL; U57911; AAC50564.1;
DR EMBL; BC031582; AAH31582.1;
DR Genew; HGNC:1180; C11orf8.
DR MIM: 600911;
DR GO; GO:0007399; P-neurogenesis; TAS.
DR InterPro; IPR004843; M-peptidase.
DR Pfam; PF00149; Metallophos; 1.
SQ SEQUENCE 294 AA; 33360 MW; 43B2BCDA1BFDF0 CRC64;

Query Match 4.5%; Score 91; DB 1; Length 294;
Best Local Similarity 20.2%; Pred. No. 17;
Matches 53; Conservative 30; Mismatches 89; Indels 90; Gaps 9;

QY 139 DIIPDEIIPRAHRLCSQPFEEVERKCRSYSHVL-----KEYRFFNLQSL 187
DB 45 DIPYDTPKPAHTRVCISDHSNRDQGMXYGDLHTGFTGLSPSEVKKFN----- 100
QY 188 YLLKDPSSLNLHVLVDRPRAVRSRERTKGLMIDSRIV-GQHEQKLKE-----D 240
DB 101 -----DMLGNLYEYKIVYAGNHELFDEKFEFADLVK 132
QY 241 QPYVMQYICQSOLEIKTQSLPKALQERYLLVREEDARPAVQTSRM-----YFVGL 296
DB 133 QDYTRPPSVSKLPEDFDVQSL-----LTNSIYQDSVYTKGFFIYA 177
QY 297 EFLPHLOTWVNIITRKSGKGDHAFITNARDALNVQAWMSLPYEKVSRLQKACGDANML 356
DB 178 PWTPEFNGFNLPRGQSLD-----KWNLIPEIDILM-THGPP----- 216
QY 357 LGYRHVRSQEOQRNLIDLST 378
DB 217 LGFRDWPVKELQRYVCVELLNT 238

RESULT 9
MX3_RAT
ID MX3_RAT STANDARD; PRT; 659 AA.
AC P18590;
DT 01-NOV-1990 (Rel. 16, Created)
DT 01-NOV-1990 (Rel. 16, Last sequence update)

```


DT 01-FEB-1995 (Rel. 31, Last annotation update)
 DE Interferon-induced GTP-binding protein Mx3.
 GN Mx3.
 OS Rattus norvegicus (Rat).
 CC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 CC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 RN NCBI_TaxID=10116;
 RX MEDLINE=91055598; PubMed=2173790;
 RA Meier E., Kunz G., Haller O., Arnhelter H.;
 RT "Activity of rat Mx proteins against a rhabdovirus";
 RL J. Virol. 64:6263-6269(1990).
 CC -1- FUNCTION: DOES NOT SHOW ACTIVITY AGAINST INFLUENZA VIRUS OR VSV;
 CC ALTHOUGH IT ONLY DIFFERS FROM Mx2 BY 8 POSITIONS.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- INDUCTION: By Interferons.
 CC -1- SIMILARITY: BELONGS TO THE DYNAMIN FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X52713; CA36937.1; -
 DR PIR: S11737; S11737.
 DR InterPro: IPR001401; Dynamln.
 DR InterPro: IPR000375; Dynamln_central.
 DR InterPro: IPR003130; GED.
 DR Pfam: PF00350; dynamln_1.
 DR Pfam: PF01031; dynamln_2; 1.
 DR Pfam: PF02212; GED; 1.
 DR PRINTS: PR00195; DYNAMIN.
 DR SMART: SM00053; DYNC; 1.
 DR SMART: SM00302; GED; 1.
 DR PROSITE: PS00410; DYNAMIN; 1.
 DR Interferon induction; GTP-binding; Multigene family.
 KW NP_BIND 75 82 GTP (POTENTIAL).
 FT NP_BIND 176 180 GTP (POTENTIAL).
 FT NP_BIND 245 248 GTP (POTENTIAL).
 FT SEQUENCE 659 AA; 74951 MW; F55D63283BC865B3 CRC64;
 SQ
 Query Match 4.58; Score 91; DB 1; Length 659;
 Best Local Similarity 20.68; Pred. No. 4.9;
 Matches 65; Conservative 50; Mismatches 104; Indels 96; Gaps 16;
 QY 62 FGHPDVFYLMPEAMHVMFPGKSTAMWLMHVAVDLI---RAVELCDMSVFDAVME--- 114
 DB 301 FKEHPORALLLEDG-----KATVPCIAERLTWELLISHICKSLPLENQIKESHQSTSE 353
 QY 115 -----PGPRRSSLFQWENSRAICAPACDIIPDEIIPRAGHRLCS-QQPF--- 161
 DB 354 ELQKYGADIDEDNEKTLFLEKINAF-NODITAIVEGEELVREKECLFTKLKEPFLM 412
 QY 162 -EVEKACRSYHVLKEVREFNQ-----SLYPLKDSLNLI 200
 DB 413 SEELERNFQKSDALYEVYEFEMOYGRRELPGFVNKTFENILIRQIKTLEBPAME-- 470
 QY 201 VHLVRD-PRAVFR-----SRRTKGLMIDSRIVMGHOKLKKEDEPYVMOV 248
 DB 471 LHKTEIVRAFTVSEKNSSEFNLRHTYTSKLEDIRL---EOTEKAKIRLHFQWEO 527
 QY 249 ICQSQLEIYKTQSLPALDERVLLVREDE-----ARAPYAQTSRYEYVGL 296
 DB 528 ILYQODIYR-----KALOK-----VREEAEEERKKGKRSASQSPMLQTSMD----- 572
 QY 297 EFLPLQTM---VHN 308
 DB 573 EIFOHLNAYROEHN 587

RESULT 10
 C7D8_SOYBN
 ID C7D8_SOYBN STANDARD: PRT: 504 AA.
 AC 081974;
 DT 15-DEC-1998 (Rel. 37, Created)
 DT 15-DEC-1998 (Rel. 37, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Cytochrome P450 71D8 (EC 1.14.-.-) (P450 CYP7).
 GN CYP71D8.
 OS Glycine max (Soybean).
 CC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
 CC Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Rosidae;
 CC eurosids I; Fabales; Fabaceae; Papilionoideae; Phaseoleae; Glycine.
 RN NCBI_TaxID=38477;
 RX MEDLINE=9648734;
 RA Schopfer C.R., Ebel J.;
 RT "Identification of elicitor-induced cytochrome P450s of soybean
 RL (Glycine max L.) using differential display of mRNA."
 CC MoI. Gen. Genet. 258:315-322(1998).
 CC -1- SIMILARITY: Belongs to the cytochrome P450 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: Y10493; CA71517.1; -
 DR PIR: T07120; T07120.
 DR HSSP: P14779; IJPZ.
 DR InterPro: IPR001128; Cytochrome_P450.
 DR Pfam: PF00067; P450; 1.
 DR PRINTS: PR00385; P450.
 DR PROSITE: PS00086; CYTOCHROME_P450; 1.
 DR Oxidoreductase; Monooxygenase; Heme.
 KW METAL 444 444 IRON (HEME AXIAL LIGAND) (BY SIMILARITY).
 FT SEQUENCE 504 AA; 57579 MW; 86E4FAEEC2FE2A8F CRC64;
 SQ
 Query Match 4.48; Score 89.5; DB 1; Length 504;
 Best Local Similarity 19.98; Pred. No. 4.7;
 Matches 70; Conservative 45; Mismatches 108; Indels 129; Gaps 16;
 QY 68 VFYLMPEAMHVMFPGKSTAMWLMHVAVDLI---RAVELCDMSVFDAVMEPGPRSSLFQWE 127
 DB 14 VFLLIH-WLV-KTYKQKSHKL-----PGPWR----- 39
 QY 128 NSRALCSAPACDIIPDEIIPRAGHRLCSQQPEVEKACRSYS---HVLKEVREFNL 184
 DB 40 -----LP---IGNLHQLAALASLPDQALQKLVKRGPLMHLQLGESTLVV 83
 QY 185 QSLYPLKDSPLNLHIVLVDPRAVFRFSRRTKGLMIDSRIVMGHOKLKKEDEPY 244
 DB 84 SS--PKAMEMKTHDHFVQRPQ-----LAPQFMYAGTDIAFADYGDYWR 129
 QY 245 VMQVYCQLEIYKTQSLPALOE---RYLLVREDELRAPVAQTSRYEYVGL----- 296
 DB 130 QIRKICTELLSARVQSFSHIRQDENKKLQSHSSAGSPIDLSGKFLSLGTTVSRRA 189
 QY 297 -----EFL-----PHLQTMVHNITRGKGMGDAFHNTA----- 324
 DB 190 FGKENDODEFMSVLRKAITWTGFEVDMPPLKP-LHLTLTRQAKAVEHV-HORADKIL 247
 QY 325 RDAIIVSQAMRMWSLPYEKVSRLQKACGAMNLGVRHVRSPQEDRNILLDL 376
 DB 248 EDILRKHM-----EKRTVKRKGNG-----SEAEQEDLVAVLL 279

RESULT 11
PKSM_BACSU STANDARD; PRT; 4273 AA.
AC P40872; 031781;
DT 01-FEB-1995 (Rel. 31, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 28-FEB-2003 (Rel. 41, Last annotation update)
DE Putative polyketide synthase pksM.
GN PKSM OR PKSY.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G., Azevedo V., Bertolo M.G., Bessières P., Bolotin A., Borchert S., Boriss R., Bourcier L., Brans A., Braun M., Brignell S.C., Bron S., Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M., Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A., Denzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T., Enlitan K.D., Erlington J., Fabret C., Ferrati E., Foulger D., Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N., Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G., Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A., Hilbert H., Holstappel S., Hosono S., Hulio M.F., Itaya M., Jones L., Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C., Kobayashi Y., Koetter P., Koningsstein G., Krog S., Kumano M., Kurita K., Lapdus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portetlelle D., Portwolk S., Prescott A.M., Prescan E., Puig P., Purnelle B., Rapoport G., Rey M., Reynolds S., Rieger M., Rivolta C., Rocha E., Roche B., Rose B., Sadaie Y., Sato T., Scanlan E., Schleich S., Schroeter R., Scofield F., Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B., Sorokin A., Taconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Yamakoshi A., Tanaka T., Terstra P., Tognoni A., Toso V., Uchiyama S., Vandendol M., Vannier F., Vassarotti A., Viari A., Wambitt R., Wedler E., Wedler H., Wetzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K., Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Danchin A., "The complete genome sequence of the Gram-positive bacterium Bacillus subtilis".
RT Nature 390:249-256(1997).
RN [2]
RP SEQUENCE OF 1-1763 FROM N.A.
RC STRAIN=168 / PBL124;
RA Tognoni A., Grandi G.;
RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN SECONDARY METABOLISM.
CC -1- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES (POTENTIAL).
CC -1- SIMILARITY: Contains 4 acyl carrier domains.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation in the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).
CC
CC EMBL; 299113; CAB13603.1;
CC EMBL; 235133; CA84505.1;
CC PIR; C69679; C69679.
CC Subtilist; BG10931; PKSM.
CC InterPro; IPR002198; ADH_short.
CC InterPro; IPR000794; Ketoacyl-synt.

DR InterPro: IPR001601; Methyltransf.
DR InterPro: IPR006163; pp_bind.
DR InterPro: IPR006162; Pantine attach.
DR InterPro: IPR000051; SAM_bind.
DR pfam: PF00106; adh_short; 1.
DR pfam: PF02801; ketoacyl-synt; 3.
DR pfam: PF00550; pp-binding; 4.
DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.
DR PROSITE: PS00606; B_KETOACYL-SYNTHASE; 2.
DR PROSITE: PS0075; ACP_DOMAIN; 4.
KW Transferase; Acyltransferase; Antibiotic biosynthesis; NADP; Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.
FT DOMAIN 295
FT DOMAIN 364
FT DOMAIN 396
FT DOMAIN 2190
FT DOMAIN 2258
FT DOMAIN 2322
FT DOMAIN 3532
FT DOMAIN 3410
FT DOMAIN 4140
FT BINDING 327
FT BINDING 2222
FT ACT_SITE 2476
FT BINDING 3446
FT ACT_SITE 3690
FT BINDING 4172
FT CONFLICT 103
FT CONFLICT 276
FT CONFLICT 289
SQ SEQUENCE 4273 AA; 477459 MW; 3BBFC1A250AEB5A CMC64;
Query Match 4.48; Score 89.5; DB 1; Length 4273;
Best Local Similarity 22.98; Pred. No. 72;
Matches 48; Conservative 34; Mismatches 87; Indels 41; Gaps 10;

QY 161 FEYVERACRSYSHVKEVRFNLQSLYPLKPS---LNLHIVLVDPRAVFSRER 216
DB 36 YQVQEHGAYQGLEIK-----NLTFPLINDESIDIALITVSRREGTSITIDGQ 90
QY 217 TKGDLMDISRIYVQGE-OKLKEDQPYVMQYICOSQLEIYKTSIPKALQERYLV 275
DB 91 QHGESLSDKR---QYETADMHRKEQTAFNEI---DLNOMSTADRLINDEIYEQOR 142
QY 276 YEDLAPAPYAQT-SRATY-----FYGLEPLHLQVWVNIIT--RKGMDAFHFN 323
DB 143 SELVHTGMMKAGQYIEKAGVAVIDLAVQGEVLRHSDAFLEHPTLIDSGIGS----- 196
QY 324 ARDALNVSOANRMSLPEKYS---RLQKAC 350
DB 197 ---SCLISDOTMTPLIYESFSASERLQKC 224

RESULT 12
FAST-FLACH STANDARD; PRT; 320 AA.
AC P52837;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-OCT-1996 (Rel. 34, Last annotation update)
DE Flavonol 4'-sulfotransferase (EC 2.8.2.-) (F4-SF).
OS Flaveria chloroefolia.
OC Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; eudicotyledons; core eudicots; Asteridae; campanulids; Asterales; Asteraceae; Tageteae; Flaveria.
OC NCBI_TaxID=4228;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=92159034; PubMed=1741382;
RA Varin L., Deluca V., Ibrahim R.K., Brisson N.;
RT "Molecular characterization of two plant flavonol sulfotransferases.";
RL Proc. Natl. Acad. Sci. U.S.A. 89:1286-1290(1992).
RN [2]

RP PAP-BINDING SITE.
 RX MEDLINE=95279378; PubMed=7759495;
 RA Varin L., Marsolais F., Brisson N.;
 RT "Chimeric flavonol sulfotransferases define a domain responsible for
 RT substrate and position specificities."
 RL J. Biol. Chem. 270:12498-12502(1995).
 CC -1- FUNCTION: TRANSFERS SULFATE GROUP INTO FLAVONOLIN POSITION 4'.
 CC MAY PLAY A ROLE IN AUXIN TRANSPORT.
 CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC.
 CC -1- TISSUE SPECIFICITY: HIGHEST IN SHOOT TIPS AND LOWEST IN MATURE
 CC LEAVES AND ROOTS (BY SIMILARITY).
 CC -1- SIMILARITY: BELONGS TO THE PLANT SULFOTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M64136; AAA33343.1; -
 DR HSSP: P50224; 1CUM;
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 DR ProDom: PD001218; Sulfotransferase; 1.
 KM Transferase.
 FT BINDING 148 163 PAPS-BINDING SITE.
 SQ SEQUENCE 320 AA; 37255 MW; 6B26359F6CCBCC0 CRC64;
 Query Match 4.4%; Score 89; DB 1; Length 320;
 Best Local Similarity 20.8%; Pred. No. 2.9;
 Matches 56; Conservative 30; Mismatches 83; Indels 100; Gaps 12;
 QY 54 GSSEVQGLRGQHP-DVFIYMEPRAMHWMTFKOSTAMLMHVAVDLIRAVFLCMSVFDAY 112
 Db 47 GALLAQSFKARDVDVFLCSYP-----KSGTTWKLALVAIVTR-----EKDER 91
 QY 113 MEP-----GPRROSSLFO-WENSALCSAPACDIIPODEIIPRACHRLCSQOPE 162
 Db 92 TSPLLNINPNCIPYEIKDKKIVENNNSCFTR-----NATHPIYH 133
 QY 163 VVEKACRSYSHVVLKEVRFNLSLYPLKDPISLNIHVLVDPRAVFRSRETKGDL 222
 Db 134 VLKRS-----IL-----ALNCKMYIYIRIKDY----- 156
 QY 223 IDRIWVGHEOKLKKEDOPYYVMQVYCOSOLEIYKIQSLP-----KALOR-- 270
 Db 157 IVSFIHGRITIKLPLEDAPFE-----EAFDEFYHGISQFGPYWHLGLGYMKASLERPE 210
 QY 271 -YLLVRYEDLARAFAVQTSRMVEFVGLF 298
 Db 211 VILFLKEDYKDKPTSVNKLAEFIGYF 239
 RESULT 13
 MPPA_RAT
 ID MPPA_RAT STANDARD: PRT; 524 AA.
 AC P20069;
 DT 01-FEB-1991 (Rel. 17, Created)
 DT 01-FEB-1991 (Rel. 17, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Mitochondrial processing peptidase alpha subunit, mitochondrial
 DE precursor (EC 3.4.24.64) (Alpha-MPP) (P-55).
 GN INPSE OR PMPCA OR MPPA.
 OS Rattus norvegicus (Rat).
 OC Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 OX NCBI_Taxid=10116;
 RN [1]
 RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
 RC STRAIN=Sprague-Dawley; TISSUE=Liver;
 RX MEDLINE=91045920; PubMed=2236012;

RA Kleiber J., Kalousek F., Swaroop M., Rosenberg L.E.;
 RT "The general mitochondrial matrix processing protease from rat liver:
 RT structural characterization of the catalytic subunit."
 RL Proc. Natl. Acad. Sci. U.S.A. 87:7978-7982(1990).
 CC -1- FUNCTION: THE MITOCHONDRIAL PROCESSING PROTEASE (MPP-I) CLEAVES
 CC PRESEQUENCES FROM MITOCHONDRIAL PROTEIN PRECURSORS. MOST MPP-I
 CC CLEAVAGE SITES FOLLOW AN ARGININE AT POSITION -2.
 CC -1- CATALYTIC ACTIVITY: Release of N-terminal transit peptides from
 CC precursor proteins imported into the mitochondrion, typically with
 CC Arg in position P2.
 CC -1- SUBUNIT: Heterodimer of alpha and beta subunits.
 CC -1- SUBCELLULAR LOCATION: Mitochondrial matrix.
 CC -1- SIMILARITY: BELONGS TO PEPTIDASE FAMILY M16.
 CC -1- CAUTION: Does not seem to have protease activity as it lacks the
 CC zinc-binding site.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL: M57728; AAA41632.1; -
 DR PIR: A36205; A36205.
 DR MEROPS: M16.971; -
 DR InterPro: IPR001431; Peptidase_M16.
 DR Pfam: PF00675; Peptidase_M16; 1.
 DR Pfam: PF05193; Peptidase_M16_C; 1.
 DR PROSITE: PS00143; INSULINASE; 1.
 KM Hydroxylase; Metalloprotease; Mitochondrion; Transit peptide.
 FT TRANSIT 1 32
 FT CHAIN 33 524
 FT MITOCHONDRIAL PROCESSING PEPTIDASE ALPHA
 SQ SEQUENCE 524 AA; 58607 MW; BBF08FBC9F09DB2 CRC64;
 Query Match 4.4%; Score 89; DB 1; Length 524;
 Best Local Similarity 21.5%; Pred. No. 5.4;
 Matches 70; Conservative 46; Mismatches 104; Indels 106; Gaps 19;
 QY 70 YLM--EPAM-----HWMTFQ-----STAMLMHVAVDLIRAVFLCMSVFDAYMEPG 116
 Db 269 YLLGVCPAMGAPGAVVMTLQWHSSTRGSSRW-----RE-----TCOMSA-----LRP 311
 QY 117 PROSSLFQWENSALCSAPACDIIPODEIIPRACHRLC-----SQOPFEVYE 165
 Db 312 PROSSH-IYGAHEL-----LLLEDFIPFAVLNMMGGGGSFAGGPKMFS--- 360
 QY 166 KACRSYSHVVLKEVRFNLSLYPLKDPISLNIHVLVDPRAVFRSRETKGDLIDS 225
 Db 361 ---RLYINLVNRHHMMNATSYHNSYED--TGLLCIHASADPRQV-----REWEIITKE 410
 QY 226 RIYVGQHEOKLKKEDOPYYVMQVYCOSOLEIYKIQSLPRALOERYLLVRYEDLARAFA 285
 Db 411 FILMGRTVDVLELERAFTQLSMLMN-----LESRP-----VIFEDVGROYLA 454
 QY 286 QTSRMVEFVGLFELPH-LQTVVNHITRGKGGDHAFTNARDALNVSQARWMSLPYEKVS 344
 Db 455 THSRK-----LPHELCTLIIRNV-----KPEDIKRVASKMLRGKP----- 488
 QY 345 RLQACGDANMLLGYRHVRSEQEQRN 370
 Db 489 -AVALGDLVDLPYEHIOALASRD 513
 RESULT 14
 CAPP_STRPY
 ID CAPP_STRPY STANDARD: PRT; 920 AA.
 AC 09A007;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)

OY 278 DLARAPVAQTSRMKEFVGLFPLPLQTW---VHNIT--RGKGM 315
Db 596 NNLYGLNNTYEGKRAVTVGVONL-----WSNVHEVSTKRHKGM 633

Search completed: August 9, 2003, 16:43:05
Job time : 12.4755 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:40:04 ; Search time 34.4109 Seconds
(without alignments)
2894.677 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLPKRMKLLFLVSQMAIL.....EQRNLLDLSTWTFVEQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues

Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPRMBL_23:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_rvivirus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Length	ID	Description
1	2038	100.0	386	4	Q9Y5R3
2	2035	99.9	386	4	Q8NCG5
3	1963	96.3	370	4	Q8IV46
4	1515.5	74.4	388	11	Q9RI11
5	1513.5	74.3	388	11	Q9WUE5
6	1019	50.0	395	4	Q9GZX3
7	1017.5	49.9	395	11	Q9QUP4
8	1008	49.5	411	4	Q9GZS9
9	627	30.8	484	11	Q9E78
10	624	30.6	484	11	Q99NB0
11	606.5	29.7	486	4	Q75667
12	605.5	29.7	486	4	Q9NS84
13	598.5	29.4	530	11	Q88276
14	591.5	29.0	483	4	Q9UEB5
15	591.5	29.0	530	4	Q9Y4C5
16	559	27.4	411	11	Q9EBC0

17	549	26.9	411	4	Q43916	Q43916 homo sapien
18	504.5	24.8	472	11	Q88199	Q88199 mus musculu
19	482	23.7	479	4	Q75099	Q75099 homo sapien
20	481.5	23.6	474	11	Q9QRL2	Q9QRL2 ratu
21	423.5	20.8	441	13	Q93403	Q93403 torped cal
22	264	13.0	486	5	Q9VMC3	Q9VMC3 drosophila
23	215	10.5	120	11	Q9DOK5	Q9DOK5 mus musculu
24	210	10.3	363	5	Q9VMC4	Q9VMC4 drosophila
25	199	9.8	363	5	Q8MZD1	Q8MZD1 drosophila
26	175	8.6	119	6	Q95Z78	Q95Z78 oryctolagus
27	168	8.2	307	16	Q92V64	Q92V64 thizobium m
28	124.5	6.1	274	16	Q8PRA0	Q8PRA0 xanthomonas
29	115.5	5.7	203	16	Q93JE6	Q93JE6 streptomyce
30	111	5.4	655	6	Q9BGC0	Q9BGC0 macaca fasc
31	108.5	5.3	1222	4	Q8IZU8	Q8IZU8 homo sapien
32	107	5.3	1586	4	Q9P2D3	Q9P2D3 homo sapien
33	100.5	4.9	747	4	Q96Q02	Q96Q02 homo sapien
34	100.5	4.9	762	17	Q8Q0G7	Q8Q0G7 methanosarc
35	100.5	4.9	775	4	Q9C011	Q9C011 homo sapien
36	100	4.9	593	10	Q8VZE0	Q8VZE0 arabidopsis
37	97.5	4.8	285	2	Q8KRX0	Q8KRX0 xanthomonas
38	96.5	4.7	743	4	Q9H0K2	Q9H0K2 homo sapien
39	96	4.7	2454	3	Q9UV56	Q9UV56 emericella
40	96	4.7	2454	3	Q9UVP2	Q9UVP2 emericella
41	95	4.7	2887	2	Q8GAC3	Q8GAC3 lynchya maj
42	93.5	4.6	529	5	Q9V490	Q9V490 drosophila
43	93.5	4.6	1201	11	Q8C7A8	Q8C7A8 mus musculu
44	93	4.6	637	11	Q8C773	Q8C773 mus musculu
45	92.3	4.5	2334	5	Q24135	Q24135 drosophila

ALIGNMENTS

RESULT 1
ID Q9Y5R3 PRELIMINARY; PRT; 386 AA.
AC Q9Y5R3;
DT 01-NOV-1999 (TREMUR12, 12, Created)
DT 01-NOV-1999 (TREMUR12, 12, Last sequence update)
DT 01-MAR-2003 (TREMUR12, 23, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (U-selectin ligand
DE sulfotransferase GSN-3).
GN GSN3.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
OX NCBI_TaxID=9606;
RX MEDLINE=99264336; PubMed=10330415;
RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstitution
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
[2]
SEQUENCE FROM N.A.
RC TISSUE=Tonsil;
RC Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
RA Rosen S.D.;
RT "Chromosomal localization and genomic organization for the
RT galactose/N-acetylgalactosamine/N-acetylglucosamine 6-O-
RT sulfotransferase gene family.";
RL Glycobiology 0:0-0(2001).
[3]
SEQUENCE FROM N.A.
RX MEDLINE=21332592; PubMed=11439191;
RA Yeh J.-C., Hiraoka N., Petryniak B., Nakayama J., Ellies L.G.,
RA Rakuda D., Hindsgraul O., March J.D., Lowe J.B., Rakuda M.;
RT "Novel sulfated lymphocyte homing receptors and their control by a
RT core1 extension beta1,3-N-acetylglucosaminyltransferase.";

RL Cell 105:957-969(2001).
 DR EMBL: AF131325; AAD3015.1; -
 DR EMBL: AF280088; AAG48246.1; -
 DR EMBL: AF149783; AAK48417.1; -
 DR InterPro: IPR001092; HLH basic.
 DR InterPro: IPR000663; Sulfoltransferase.
 DR Pfam: PF00685; Sulfoltransferase.
 DR PROSITE: PS00038; HLH_1; 1.
 DR Lectin; Selectin; Transferase.
 KW SEQUENCE 386 AA; 45133 MW; 0C3BB402241743A CRC64;

Query Match 100.0%; Score 2038; DB 4; Length 386;
 Best Local Similarity 100.0%; Pred. No. 7,4e-184;
 Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLTPKMKLLFLVSGMATALEFFHMYSHNISLSKKAQPERMHVYLSSWSSGSSFFVQ 60
 Db 1 MLTPKMKLLFLVSGMATALEFFHMYSHNISLSKKAQPERMHVYLSSWSSGSSFFVQ 60
 QY 61 LFGQHPDVYLMERPAHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 Db 61 LFGQHPDVYLMERPAHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 QY 121 SSLEFQENSRLACSPACDIIPODEIIPRAHCRLLCSQOPFEVEKACRSYSHVYLKEVR 180
 Db 121 SSLEFQENSRLACSPACDIIPODEIIPRAHCRLLCSQOPFEVEKACRSYSHVYLKEVR 180
 QY 181 FFMQSLYPLKDPISLHIVLVRDPAVFSRERTKGDMLDSRIYVGOHEQKLKED 240
 Db 181 FFMQSLYPLKDPISLHIVLVRDPAVFSRERTKGDMLDSRIYVGOHEQKLKED 240
 QY 241 OPIYVMOVICOSOLEYTKTOSLPKALOERYLLVREDELRAPVQTSRYEFGLEFLP 300
 Db 241 OPIYVMOVICOSOLEYTKTOSLPKALOERYLLVREDELRAPVQTSRYEFGLEFLP 300
 QY 301 HLOTWVHNTTRGKMGDHAFTNARDALNVSQAMRSLPEYKYSRLQKACGDMNLGGR 360
 Db 301 HLOTWVHNTTRGKMGDHAFTNARDALNVSQAMRSLPEYKYSRLQKACGDMNLGGR 360
 QY 361 HVRSQEQRNLLDLSTVTPEQIH 386
 Db 361 HVRSQEQRNLLDLSTVTPEQIH 386

RESULT 2

QY 08NCG5 PRELIMINARY; PRT; 386 AA.
 Db 08NCG5;

DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-OCT-2002 (Tremblrel. 22, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Hypothetical protein FLJ90265.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
 RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
 RA Yamamoto J., Nakamatsu A., Nakamura Y., Kojima S., Nagahara K.,
 RA Masuko Y., Ono T., Okeno K., Yoshikawa Y., Aotsuka S., Sasaki N.,
 RA Hattori A., Okumura K., Iwayanagi T., Ninomiya K.;
 RT "NEO human cDNA sequencing project."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK074746; BAC1177.1; -
 DR Gene: HGNC:1972; CHST4.
 KW Hypothetical protein.
 SQ SEQUENCE 386 AA; 45160 MW; 861869348319E42A CRC64;

Query Match 99.9%; Score 2035; DB 4; Length 386;
 Best Local Similarity 99.7%; Pred. No. 1.4e-183;
 Matches 385; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLTPKMKLLFLVSGMATALEFFHMYSHNISLSKKAQPERMHVYLSSWSSGSSFFVQ 60
 Db 1 MLTPKMKLLFLVSGMATALEFFHMYSHNISLSKKAQPERMHVYLSSWSSGSSFFVQ 60
 QY 61 LFGQHPDVYLMERPAHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 Db 61 LFGQHPDVYLMERPAHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQ 120
 QY 121 SSLEFQENSRLACSPACDIIPODEIIPRAHCRLLCSQOPFEVEKACRSYSHVYLKEVR 180
 Db 121 SSLEFQENSRLACSPACDIIPODEIIPRAHCRLLCSQOPFEVEKACRSYSHVYLKEVR 180
 QY 181 FFMQSLYPLKDPISLHIVLVRDPAVFSRERTKGDMLDSRIYVGOHEQKLKED 240
 Db 181 FFMQSLYPLKDPISLHIVLVRDPAVFSRERTKGDMLDSRIYVGOHEQKLKED 240
 QY 241 OPIYVMOVICOSOLEYTKTOSLPKALOERYLLVREDELRAPVQTSRYEFGLEFLP 300
 Db 241 OPIYVMOVICOSOLEYTKTOSLPKALOERYLLVREDELRAPVQTSRYEFGLEFLP 300
 QY 301 HLOTWVHNTTRGKMGDHAFTNARDALNVSQAMRSLPEYKYSRLQKACGDMNLGGR 360
 Db 301 HLOTWVHNTTRGKMGDHAFTNARDALNVSQAMRSLPEYKYSRLQKACGDMNLGGR 360
 QY 361 HVRSQEQRNLLDLSTVTPEQIH 386
 Db 361 HVRSQEQRNLLDLSTVTPEQIH 386

RESULT 3

QY 08IV46 PRELIMINARY; PRT; 370 AA.
 Db 08IV46;

DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Similar to carbohydrate (N-acetylglucosamine 6-O) sulfoltransferase
 4.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tissue-Brain;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC035282; AAH35282.1; -
 KW Transferase.
 SQ SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;

Query Match 96.3%; Score 1663; DB 4; Length 370;
 Best Local Similarity 100.0%; Pred. No. 8.4e-177;
 Matches 370; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 17 MALALFFHMYSHNISLSKKAQPERMHVYLSSWSSGSSFFVQGLFGQHPDVYLMERPAW 76
 Db 1 MALALFFHMYSHNISLSKKAQPERMHVYLSSWSSGSSFFVQGLFGQHPDVYLMERPAW 76
 QY 77 HMYTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQSSLEFQENSRLACSP 136
 Db 77 HMYTFKOSTAMMLHMAVRDLIRAVFLCDMSVFDAYMEPGRQSSLEFQENSRLACSP 136
 QY 137 ACDDIPODEIIPRAHCRLLCSQOPFEVEKACRSYSHVYLKEVRFNLOSILPLKDPSTL 196
 Db 137 ACDDIPODEIIPRAHCRLLCSQOPFEVEKACRSYSHVYLKEVRFNLOSILPLKDPSTL 196
 QY 197 NHIYHVLVDPRAVFSRERTKGDMLDSRIYVGOHEQKLKEDQPIYVMOVICOSOLEI 256
 Db 197 NHIYHVLVDPRAVFSRERTKGDMLDSRIYVGOHEQKLKEDQPIYVMOVICOSOLEI 256
 QY 257 YTTIOSLPKALOERYLLVREDELRAPVQTSRYEFGLEFLPHLOTWVHNTTRGKMG 316
 Db 257 YTTIOSLPKALOERYLLVREDELRAPVQTSRYEFGLEFLPHLOTWVHNTTRGKMG 316

```

Db      241 YKTQSLPKALQERYLLVREYEDLARAPVAQTSRMVEFGLFEPLQTVHNITRGKGMG 300
Oy      317 DHAFHTNARDALNTNSQAWRMSLPEKYSRLCKACGDANMLGYRHVRSEQRNLLD 376
Db      301 DHAFHTNARDALNTNSQAWRMSLPEKYSRLCKACGDANMLGYRHVRSEQRNLLD 360
Oy      377 STWTVPEQIH 386
Db      361 STWTVPEQIH 370

```

RESULT 4

```

O9RL11 PRELIMINARY: PRT: 388 AA.
ID 09RL11
AC 09RL11
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE L-selectin ligand sulfotransferase.
GN CHST4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=99361934; PubMed=10435581;
RA Hiraoka N., Petyrniak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,
RA Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.;
RT "A novel, high endothelial venule-specific sulfotransferase expresses
RT 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34.";
RL Immunity 11:79-89(1999).
DR EMBL: AF109155; AAd45579.1;
DR MGD: MGI:1349479; Chst4.
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam: PF00685; Sulfotransfer; 1.
DR Lectin: Selectin; Transferrase.
SQ SEQUENCE 388 AA; 44636 MW; 6D5371AFB6884AEE CRC64;

```

```

Query Match 74.4%; Score 1515.5; DB 11; Length 388;
Best Local Similarity 72.8%; Pred. No. 1.5e-134;
Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;

Oy      1 MLPRKMKLLFLVSQMAIILFFHMYSHNITSLSMKQAPR-MHVLVLSWRSGSSPVG 59
Db      1 MMLKKGRLMFLSQVIVLALFTIMSVHR--HLSQRESRPPVHVLVLSWRSGSSPVG 58
Oy      60 QLFQGHDPVFLMEPAMHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVPFAYMEPPRR 119
Db      59 QLFQGHDPVFLMEPAMHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVPFAYMPPRR 118
Oy      120 QSSLFQWENSALCSAPACDIIPODEIPRAHCRLLCSQGFVEVEKACRSYSHVYLKEV 179
Db      119 QSSLFQWENSALCSAPACDIIPODEIPRAHCRLLCSQGFVEVEKACRSYSHVYLKEV 178
Oy      180 RFFNLQSLYPLLKDPSLNLHIVHLVRODRAVFRSREKGLMDISRTVMGQHEOKLKE 239
Db      179 RFLSLQALYPLLKDPSLNLHIVHLVRODRAVFRSREKGLMDISRTVMGQHEOKLKE 238
Oy      240 DQPYVMQVITQSOLEIKYKTQSLPKALQERYLLVREYEDLARAPVAQTSRMVEFGLFE 299
Db      239 DQPYVMQVITQSOLEIKYKTQSLPKALQERYLLVREYEDLARAPVAQTSRMVEFGLFE 298
Oy      300 PHLTQVHNITRGKGMGHAFHTNARDALNTNSQAWRMSLPEKYSRLCKACGDANMLGY 359
Db      299 PHLTQVHNITRGKGMGHAFHTNARDALNTNSQAWRMSLPEKYSRLCKACGDANMLGY 358
Oy      360 RHVRSEQRNLLDLSSTWTVPEQI 385
Db      359 LQVRSQEQEGNLSLIDLSSSHLIGV 384

```

```

RESULT 5
O9WDE5 PRELIMINARY: PRT: 388 AA.
ID 09WDE5
AC 09WDE5
DT 01-NOV-1999 (TREMBLrel. 12, Created)
DT 01-NOV-1999 (TREMBLrel. 12, Last sequence update)
DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
DE N-acetylglucosamine 6-O-sulfotransferase (carbohydrate (Chondroitin
DE 6/keratan) sulfotransferase 4).
GN CHST4.
OS Mus musculus (Mouse).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX NCBI_TaxID=10090;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6; TISSUE=Tonsil;
RX MEDLINE=99264336; PubMed=10330415;
RA Bistrup A., Bhakta S., Lee J.-K., Belov Y.Y., Gunn M.D., Zuo F.-R.,
RA Huang C.-C., Kannagi R., Rosen S.D., Hemmerich S.;
RT "Sulfotransferases of two specificities function in the reconstruction
RT of high endothelial cell ligands for L-selectin.";
RL J. Cell Biol. 145:899-910(1999).
RN [2]
RP SEQUENCE FROM N.A.
RX STRAIN=C57BL/6J; TISSUE=Tongue;
RX MEDLINE=21085660; PubMed=11217851;
RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,
RA Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,
RA Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,
RA Saito T., Okazaki Y., Gotohori T., Bono H., Kasukawa T., Saito R.,
RA Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T.,
RA Fleischmann W., Gaasterland T., Gissi C., King B., Kochava H.,
RA Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,
RA Schriml L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T.,
RA Blake J., Boffelli D., Fujuno M., Aono H., Baldarelli R., Barsh G.,
RA Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,
RA Gustigich S., Hill D., Hofmann M., Hume D.A., Kanlaya M., Lee N.H.,
RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
RA Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
RA Sasaki H., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.-F.,
RA Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilmink L.,
RA Wynshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S.,
RA Hayashizaki Y.;
RT "Functional annotation of a full-length mouse cdna collection.";
RL Nature 409:685-690(2001).
DR EMBL: AF131236; AAD33016.1;
DR EMBL: AK009113; BAB36078.1;
DR MGD: MGI:1349479; Chst4.
DR InterPro: IPR000863; Sulfotransferase.
DR Pfam: PF00685; Sulfotransfer; 1.
DR Transferrase.
SQ SEQUENCE 388 AA; 44694 MW; D1E9D796DF8574D CRC64;

```

```

Query Match 74.3%; Score 1513.5; DB 11; Length 388;
Best Local Similarity 72.8%; Pred. No. 2.4e-134;
Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;

Oy      1 MLPRKMKLLFLVSQMAIILFFHMYSHNITSLSMKQAPR-MHVLVLSWRSGSSPVG 59
Db      1 MMLKKGRLMFLSQVIVLALFTIMSVHR--HLSQRESRPPVHVLVLSWRSGSSPVG 58
Oy      60 QLFQGHDPVFLMEPAMHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVPFAYMEPPRR 119
Db      59 QLFQGHDPVFLMEPAMHVMWTFKOSTAMMLHMAVRDLIRAVFLCDMSVPFAYMPPRR 118
Oy      120 QSSLFQWENSALCSAPACDIIPODEIPRAHCRLLCSQGFVEVEKACRSYSHVYLKEV 179
Db      119 QSSLFQWENSALCSAPACDIIPODEIPRAHCRLLCSQGFVEVEKACRSYSHVYLKEV 178
Oy      180 RFFNLQSLYPLLKDPSLNLHIVHLVRODRAVFRSREKGLMDISRTVMGQHEOKLKE 239
Db      179 RFLSLQALYPLLKDPSLNLHIVHLVRODRAVFRSREKGLMDISRTVMGQHEOKLKE 238

```


DB 179 RELSLQALPLTDLPSLNLHVHLVDRPAVRSRSHHTITELMVDSHIYLGHLFTKEE 238
 QY 240 DDPYVMQVICOQSOLEIYKTIOSLPKALOEYLLVREEDLAPAPVQTSRMTEFVGLLEP 299
 DB 239 DDPYAMKTIKCSQVQDIVAKICTLPALQORLYLEEDLVAPLQOTRLXYFVGIDFL 298
 QY 300 PHLOTVNHTITRGKMG--DHAFTNARDALNVSAQWMSLPEYKVSRLQKACGDAMNLL 359
 DB 299 PHLOTVNHTITRGKMG--DHAFTNARDALNVSAQWMSLPEYKVSRLQKACGDAMNLL 358
 QY 360 RHVRSEOEQRNLIDLSTWVPEQI 385
 DB 359 LGVRSQOEOGNLSIDLSSSHILGQV 384

RESULT 6
 Q9GZX3 PRELIMINARY; PRT; 395 AA.

Q9GZX3: PRELIMINARY; PRT; 395 AA.
 DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase GST-4beta (Corneal
 GN GST4BETA OR CHST6.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
 OX NCBI_TaxID=9606;
 RN (1)
 RP SEQUENCE FROM N.A.
 RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,
 RT "Chromosomal Localization and Genomic Organization for the
 RT Galactose/N-acetylglactosamine/N-acetylglucosamine 6-O-
 RL Glycobiology 0:0-0(2001).
 RN (2)
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20472330; PubMed=11017086;
 RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Fujiwara T.,
 RA Nakamura T., Doka A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S.,
 RA Ozaki K., Kinoshita S., Shimomura Y., Tanigami A., Fukuda M.N.;
 RT "Macular corneal dystrophy type I and type II are caused by distinct
 RT mutations in a new sulfotransferase gene."
 RL Nat. Genet. 26:237-241(2000).
 DR EMBL: AF280086; AAG48244.1; -;
 DR EMBL: AF219990; AAG26325.1; -;
 DR EMBL: AF219991; AAG26327.1; -;
 DR Gene: HGNC:6938; CHST6
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 DR KW
 SQ SEQUENCE 395 AA; 44098 MW; 433CA60248A48B67 CRC64;

Query Match 50.0%; Score 1019; DB 4; Length 395;
 Best Local Similarity 54.5%; Pred. No. 1.1e-87;

Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY 2 LPLKKMLLFLVSQMAIILAFHMYSHNITSSMKAOPERMHVILVSSMRSGSSV 61
 DB 14 LLLAQFLFLFLYSR-----GSSPRAGGARVAVLVLSMRSGSSV 59
 QY 62 FQOHDPVFLMEPAMHVMTFKOSTAMLMHMAVRDLIRAVFLCDMSVFDAYMEPGR 121
 DB 60 FQOHDPVFLMEPAMHVMTFKOSTAMLMHMAVRDLIRAVFLCDMSVFDAYMEPGR 118
 QY 122 SIFQWNSRALSAPACDIIPODEIIPRAHCRILCSQOPFEYVEKACRSYSHVYLKE 181
 DB 119 SIFQWNSRALSAPACDIIPODEIIPRAHCRILCSQOPFEYVEKACRSYSHVYLKE 178
 QY 182 FNLQSLYPLLPKPSLNLHVHLVDRPAVRSRSHHTITELMVDSHIYLGHLFTKEE 241
 DB 182 FNLQSLYPLLPKPSLNLHVHLVDRPAVRSRSHHTITELMVDSHIYLGHLFTKEE 241

DB 179 FNLQSLYPLLPKPSLNLHVHLVDRPAVRSRSHHTITELMVDSHIYLGHLFTKEE 237
 QY 242 PYVMQVICOQSOLEIYK--TIOSLPKALOEYLLVREEDLAPAPVQTSRMTEFVGLLEP 299
 DB 238 GLNVRVRCRSHVRINAEATLKP--PFLGRRLVREEDLAPAPVQTSRMTEFVGLLEP 296
 QY 300 PHLOTVNHTITRGKMG--DHAFTNARDALNVSAQWMSLPEYKVSRLQKACGDAMNLL 357
 DB 297 POLEAMHHTITRGSGGARREAFKISSRNALNVSAQWMSLPEYKVSRLQKACGDAMNLL 356
 QY 358 GYRVRSEOEQRNLIDLSTWVPEQI 386
 DB 357 LGVRSQOEOGNLSIDLSSSHILGQV 384

RESULT 7
 Q9QUP4 PRELIMINARY; PRT; 395 AA.

Q9QUP4: PRELIMINARY; PRT; 395 AA.
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase.
 GN CHST5 OR I-GLCNAC-6-ST.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6; TISSUE=Intestine;
 RX MEDLINE=99423499; PubMed=10491328;
 RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.,
 RT "Cloning and Characterization of a Mammalian N-acetylglucosamine-6-
 RT Sulfotransferase that is Highly Restricted to Intestinal Tissue."
 RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
 DR EMBL: AF176841; AAD56003.1; -;
 DR EMBL: AF176840; AAD56002.1; -;
 DR MGI: 1931825; Chst5.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 DR KW
 SQ SEQUENCE 395 AA; 44537 MW; 3FDF71843ED383BE CRC64;

Query Match 49.9%; Score 1017.5; DB 11; Length 395;
 Best Local Similarity 56.4%; Pred. No. 1.6e-87;

Matches 215; Conservative 43; Mismatches 112; Indels 11; Gaps 7;

QY 1 MLPLK--KKMLLFLVSQMAIILAFHMYSHNITSSMKAOPERMHVILVSSMRSGSSV 58
 DB 1 MLPLK--KKMLLFLVSQMAIILAFHMYSHNITSSMKAOPERMHVILVSSMRSGSSV 56
 QY 59 GOLFGQHPDVFYLMPEAMHVMTFKOSTAMLMHMAVRDLIRAVFLCDMSVFDAYMEPGR 118
 DB 57 GOLFGQHPDVFYLMPEAMHVMTFKOSTAMLMHMAVRDLIRAVFLCDMSVFDAYMEPGR 115
 QY 119 ROSSLFQWNSRALSAPACDIIPODEIIPRAHCRILCSQOPFEYVEKACRSYSHVYLKE 178
 DB 116 ROSSLFQWNSRALSAPACDIIPODEIIPRAHCRILCSQOPFEYVEKACRSYSHVYLKE 175
 QY 179 VREFNLQSLYPLLPKPSLNLHVHLVDRPAVRSRSHHTITELMVDSHIYLGHLFTKEE 238
 DB 176 VREFNLQSLYPLLPKPSLNLHVHLVDRPAVRSRSHHTITELMVDSHIYLGHLFTKEE 234
 QY 239 EDQPYVMQVICOQSOLEIYK--TIOSLPKALOEYLLVREEDLAPAPVQTSRMTEFVGLLE 297
 DB 235 EDQPYVMQVICOQSOLEIYK--TIOSLPKALOEYLLVREEDLAPAPVQTSRMTEFVGLLE 294
 QY 298 FLPHLOTVNHTITRGKMG--DHAFTNARDALNVSAQWMSLPEYKVSRLQKACGDAMNLL 355
 DB 295 FLPHLOTVNHTITRGKMG--DHAFTNARDALNVSAQWMSLPEYKVSRLQKACGDAMNLL 354
 QY 356 LGYRVRSEOEQRNLIDLSTWVPEQI 386
 DB 357 LGYRVRSEOEQRNLIDLSTWVPEQI 384

Db 355 LIGRYSVSELEQDLSIDL 375

RESULT 8

09GZS9 PRELIMINARY; PRT; 411 AA.

AC 09GZS9; 09UBY3;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Intestinal glucan-6-sulfotransferase (Intestinal N-acetylglucosamine-6-O-sulfotransferase) (N-acetylglucosamine 6-O-sulfotransferase).
 GN CHST5 OR I-GCNC-6-ST.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NC NCB1_TaxID=9606;
 [1]
 RP SEQUENCE FROM N.A.
 RA MEDLINE-20472330; PubMed-11017086;
 RA Akema T.O., Nishida K., Nakayama J., Watanabe H., Ozaki K., Nakamura T., Dora A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S., Fujiwara T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A., Fukuda M.N.;
 RT "Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene.";
 RL Nat. Genet. 26:237-241(2000).
 [2]
 RP SEQUENCE OF 22-411 FROM N.A.
 RC TISSUE=Intestine;
 RX MEDLINE-99423499; PubMed-10491328;
 RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
 RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-Sulfotransferase that is Highly Restricted to Intestinal Tissue.";
 RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
 DR EMBL; AF246718; AAC28023.1; -
 DR EMBL; AF219991; AAC26326.1; -
 DR EMBL; AF176839; AAD56001.1; -
 DR EMBL; AF176838; AAD56000.1; -
 DR Genew; HGNC:1973; CHST5.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 KW Transferrase.
 SQ SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;

Query Match 49.5%; Score 1008; DB 4; Length 411;
 Best Local Similarity 52.3%; Pred. No. 1.3e-86;
 Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

1 MLPRKKMLLEFVSOMAILALFFHMYSHNISLSMKAKQPRMHYLVLSNRSGSPFGO 60
 35 LLLAQITCLLFIITSRP-----GPPSPAGEEDRVHVLVLSNRSGSPFGO 80
 61 LFGHPDVFYLMERPAHVMTEFKOSTANMLHNAVDLIRAVELCDMSYFDAYMEGPRRO 120
 81 LFSQHPDVFYLMERPAHVMTEFKOSTANMLHNAVDLIRAVELCDMSYFDAYMEGPRRO 139
 121 SLSQWENSRALCSAPACDIIIPDEIIPRAHCRLLCSQPEVEYKAKRSRSHVLEKVR 180
 140 SAFFWATSRALCSAPACDIIIPDEIIPRAHCRLLCSQPEVEYKAKRSRSHVLEKVR 199
 181 FENIQLSLPLKDESLNLIHVLVDPRAVFRSRETRGDMIDRSIYMGHGEOLKRED 240
 200 FENIQLSLPLKDESLNLIHVLVDPRAVFRSRETRGDMIDRSIYMGHGEOLKRED 258
 241 QPYVYMOYICQSQLEIYK--TIQSLPKALQERLYLLVREYEDLARAFAVQTSRYEVEGLEF 298
 259 PHLRLIREVCRSHVIAEAATLKP-PPLRGGRYRLVREDLAREPLAELIRALVATGTTL 317
 299 LPHLOTWYHNTTRGKMD--HAFITNARDALNVSQAKRWSLPYKYSRLQACDANL 356
 318 TPQLEAWIHNTHGSGIKPIEAFTSSRNARNSOAMRHALLPFTKILRVOEVCAGALOL 377

QY 357 LGYRHVRSEQRNLLIDL-----STWTYPE 383
 Db 378 LGYRVSADQQRDLTDLVLRGPDHFWASPD 411

RESULT 9

09EP78 PRELIMINARY; PRT; 484 AA.

AC 09EP78;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase GST-5 (N-acetylglucosamine-6-O-sulfotransferase) (2600013M07Rik protein) (Carbohydrate O-sulfotransferase).
 GN CHST7 OR GST5 OR 2600013M07Rik.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NC NCB1_TaxID=10090;
 [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=C57BL/6;
 RX PubMed-10956661;
 RA Bhakta S., Bartes A., Bowman K.G., Kao M.M., Polsky I., Lee J.-K., Cook B.N., Bruhl R.E., Rosen S.D., Bertozzi C.R., Hemmerich S.;
 RT "Sulfation of N-Acetylglucosamine by Chondroitin 6-Sulfotransferase 2 (GST-5).";
 RL J. Biol. Chem. 275:40226-40234(2000).
 [2]
 RP SEQUENCE FROM N.A.
 RA Uchiyama K., Muramatsu T.;
 RT "Identification and molecular characterization of a cDNA encoding a novel N-acetylglucosamine-6-O-sulfotransferase.";
 RL Submitted (MAR-2000) to the EMBL/Genbank/DBJ databases.
 [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C57BL/6J; TISSUE=Embryo;
 RX MEDLINE-21085660; PubMed-11217851;
 RA Kawai J., Shinagawa A., Shibata K., Yoshino M., Toh M., Ishii Y., Aikawa T., Hara A., Fukunishi Y., Kono H., Aachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana I., Saito T., Okazaki Y., Gojibori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batilov S., Casavant T., Fleischmann W., Gaasterland T., Gissi C., King B., Kochia H., Kuehl P., Lewis S., Matsuo Y., Nikaio I., Pesole G., Quackenbush J., Schriml L.M., Staibill F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barsh G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Gustincich S., Hill D., Hofmann M., Hume D.A., Kamya M., Lee N.H., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombers P., Nardone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Saeki K., Sato K., Schoenbach C., Seya T., Shibata Y., Storch K.F., Suzuki H., Toyooka K., Wang K.H., Welter C., Whitaker C., Wilmink L., Wyshynski-Boris A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S., Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Kidney;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/Genbank/DBJ databases.
 DR EMBL; AF280089; AAG48247.1; -
 DR EMBL; AB040710; BAB3769.1; -
 DR EMBL; AK011202; BAB37465.1; -
 DR EMBL; BC019204; AAH19204.1; -
 DR MGI; MGI:189167; Chst7.
 DR InterPro; IPR000863; Sulfotransferase.
 DR Pfam; PF00685; Sulfotransfer; 1.
 KW Transferrase.

SQ SEQUENCE 484 AA: 54766 MW: 98195537D7AB7193 CRC64;
 Query Match 30.8%; Score 627; DB 11; Length 484;
 Best Local Similarity 34.4%; Pred. No. 1.5e-50;
 Matches 137; Conservative 65; Mismatches 150; Indels 46; Gaps 7;
 QY 30 NISSLSMKAPREMHVLTSSWRSSSFVGFQGFDPVYLMERPAHVMYMTKOSTAM 89
 DB 87 NISAVEAVTQEKQHYHATWRTGSSFLGELFNQHPDVYLYEPMMHLMQALYPGDAES 146
 QY 90 LHMVADLIRAVFLCDMSVFDAYMEGPRRQ-----SSLFQWENSRLCSAPACD 139
 DB 147 LGALALDMLRSLFRCDFSVLRQAQDGERAPDANLTTALFMRKRYKVCSPPLCP 206
 QY 140 IIPQDE----IIPRACHRLCSQOPEVEVEKACRSYSHVYLKEVRFNLSYPLKDP 195
 DB 207 AAPRARDVGLDEKACESTCPVSLRALEACRKYPPVVIKDYRLDLGLVPLRDG 266
 QY 196 LNHVHILVDRPAVRSPERTKGLMDS-----RLVM-----GQHE 233
 DB 267 LNLKVVQLFRDPRAVNSRLKSRQGLLRISIOVLRQRGDHFRHVLHAGVDARPGQA 326
 QY 234 OKLKEDOPY-----VMQYICOSLEIYKTIOGLPALOERYLLVVEDLARAPVQTSR 289
 DB 327 RALPSAPRADFLTSALVEICEAWLMDLFTRGAPAMLRRLRYLRLEEDLVMPQQLRR 386
 QY 290 MREYVGLLEPLQLTWNHNTKRGKG-DHAFTNARDALNVSQAWMSLPEKVSRLK 348
 DB 387 LRFSGLRLLAALDAFAFMRTGSAVGADRPFLHSARDAREAVHMRERLSQGVQVET 446
 QY 349 ACGDANMLGVRHVRSEORNL-----LDLISTTV 381
 DB 447 ACAPARLLAYPRSGDERKRTVREGETPLETKANAV 484
 RESULT 10
 Q99NB0 PRELIMINARY; PRT; 484 AA.
 AC Q99NB0; 01-JUN-2001 (Tremblrel. 17, Created)
 DT 01-JUN-2001 (Tremblrel. 17, last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
 DE Chondroitin 6-sulfotransferase-2.
 GN CHST7 OR MC6ST-2.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
 NCBI_Taxid=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Mammary gland;
 RA Kitagawa H., Uyama T., Sugahara K.;
 RT Cloning and Expression of Mouse Chondroitin 6-sulfotransferase-2.
 RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AB046929; BAB40372.1;
 DR MGI:1891767; Chst7
 DR InterPro: IPR000663; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Transferrase.
 SQ SEQUENCE 484 AA: 54751 MW: 402C5E1ED185FDF8 CRC64;
 Query Match 30.6%; Score 624; DB 11; Length 484;
 Best Local Similarity 34.2%; Pred. No. 2.8e-50;
 Matches 136; Conservative 66; Mismatches 150; Indels 46; Gaps 7;
 QY 30 NISSLSMKAPREMHVLTSSWRSSSFVGFQGFDPVYLMERPAHVMYMTKOSTAM 89
 DB 87 NISAVEAVTQEKQHYHATWRTGSSFLGELFNQHPDVYLYEPMMHLMQALYPGDAES 146
 QY 90 LHMVADLIRAVFLCDMSVFDAYMEGPRRQ-----SSLFQWENSRLCSAPACD 139
 DB 147 LGALALDMLRSLFRCDFSVLRQAQDGERAPDANLTTALFMRKRYKVCSPPLCP 206

QY 140 IIPQDE----IIPRACHRLCSQOPEVEVEKACRSYSHVYLKEVRFNLSYPLKDP 195
 DB 207 AAPRARDVGLDEKACESTCPVSLRALEACRKYPPVVIKDYRLDLGLVPLRDG 266
 QY 196 LNHVHILVDRPAVRSPERTKGLMDS-----RLVM-----GQHE 233
 DB 267 LNLKVVQLFRDPRAVNSRLKSRQGLLRISIOVLRQRGDHFRHVLHAGVDARPGQA 326
 QY 234 OKLKEDOPY-----VMQYICOSLEIYKTIOGLPALOERYLLVVEDLARAPVQTSR 289
 DB 327 RALPSAPRADFLTSALVEICEAWLMDLFTRGAPAMLRRLRYLRLEEDLVMPQQLRR 386
 QY 290 MREYVGLLEPLQLTWNHNTKRGKG-DHAFTNARDALNVSQAWMSLPEKVSRLK 348
 DB 387 LRFSGLRLLAALDAFAFMRTGSAVGADRPFLHSARDAREAVHMRERLSQGVQVET 446
 QY 349 ACGDANMLGVRHVRSEORNL-----LDLISTTV 381
 DB 447 ACAPARLLAYPRSGDERKRTVREGETPLETKANAV 484
 RESULT 11
 Q75667 PRELIMINARY; PRT; 486 AA.
 AC Q75667; 01-NOV-1998 (Tremblrel. 08, Created)
 DT 01-NOV-1998 (Tremblrel. 08, last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, last annotation update)
 DE D071116.4 (N-acetylglucosamine-6-O-sulfotransferase).
 GN D071116.4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Graham D.;
 RL Submitted (JUL-1998) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RA Uchimura K., Murata T.;
 RT Identification and molecular characterization of a cDNA encoding a
 RT novel N-acetylglucosamine-6-O-sulfotransferase.
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL; AL022165; CA18154.1;
 DR EMBL; AB040711; BAB13770.1;
 DR GeneW; HGNC:13817; CHST7
 DR InterPro: IPR000663; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Transferrase.
 SQ SEQUENCE 486 AA: 54266 MW: 3F1FD1430B3C8B95 CRC64;
 Query Match 29.8%; Score 606.5; DB 4; Length 486;
 Best Local Similarity 37.0%; Pred. No. 1.3e-48;
 Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;
 QY 41 ERMHVLTSSWRSSSFVGFQGFDPVYLMERPAHVMYMTKOSTAMLHNAVVDLRA 100
 DB 100 EKQHYHATWRTGSSFLGELFNQHPDVYLYEPMMHLMQALYPGDAESLGALDMLRS 159
 QY 101 VFLCDMSVFDAYMEG-PRRQ-----SSLFQWENSRLCSAPACDIIPQDE---I 146
 DB 160 LFRCDFSVLRQAQDGERAPADYANLTTALFMRKRYKVCSPPLCPGAPARAEGVL 219
 QY 147 IPRACHRLCSQOPEVEVEKACRSYSHVYLKEVRFNLSYPLKDPSTLNHIVHLYRD 206
 DB 220 VEDTACRSCPPVAIRALEACRKYPPVVIKDYRLDLGLVPLRDGLNLKVVQLFRD 279
 QY 207 PAVRSPERTKGLMDS-----RLVM-----GQHEOKLKEDOPY 244
 DB 280 PRAVINSRLKSRQGLLRISIOVLRQRGDHFRHVLHAGVDARPGQSRALPAPRADF 339

QY 245 ----VMOVICOSQLEIYKTIQSLPKALQERYLLVREYEDLARAPVACTSRMYEFVGLFELP 300
 Db 340 FLTGALVEICEMALRDLRFARCAPAWLRRLRYLRFEDLVROPRADRLRLRRSGLRALA 399
 QY 301 HLOTWVHNITRGKMG-DHAFTNARDALNVSOAMRMSLEPEVSRLOKACGDAMNLLGY 359
 Db 400 ALDAFALNMTRGAAAGADREFHLSARDAREAVHAMRELSREQVQVEVACAPAMRLAY 459
 QY 360 RHVRSQOE 367
 Db 460 --PRSGEE 465

RESULT 12

Q9NS84 PRELIMINARY; PRT; 486 AA.
 AC Q9NS84; 01-OCT-2000 (Tremblrel, 15, Created)
 DT 01-OCT-2000 (Tremblrel, 15, Last sequence update)
 DE 01-MAR-2003 (Tremblrel, 23, Last annotation update)
 GN Chondroitin 6-sulfotransferase-2.
 OS Homo sapiens (human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 NCBI_TaxID=9606;
 [1] NCBI_TaxID=9606;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Brain;
 RX MEDLINE=20347214; PubMed=10781596;
 RA Kitagawa H., Fujita M., Ito N., Sugahara K.;
 RT "Molecular cloning and expression of a novel chondroitin 6-O-
 RT sulfotransferase.";
 RL J. Biol. Chem. 275:21075-21080(2000).
 DR EMBL: AB037187; BAA03217.1; -;
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Sulfotransferase.
 SQ SEQUENCE 486 AA; 54294 MW; B524D1430B3C9075 CRC64;

Query Match

Best Local Similarity 29.7%; Score 605.5; DB 4; Length 486;
 Best Local Similarity 37.0%; Pred. No. 1.6e-48;
 Matches 136; Conservative 56; Mismatches 133; Indels 43; Gaps 8;

QY 41 ERNHVVLSSMRSGSFFVQLFGQHPDVFYIMEPAHVMTFKOSTAMMLHMAVRDLIRA 100
 Db 100 EKHIYVHNMTRGSSFFGLFNOHPDVFYIMEPAHVMTFKOSTAMMLHMAVRDLIRA 139
 QY 101 VFLCDMSVFDAYMEPG-PRRQ-----SLEFQENSRAICAPACDIIPDE---1 146
 Db 160 LFRCDPSVLTALYRPGDPAARAPDANLTLTAALFRWNTNKVICSPLCPAPARAREVGL 219
 QY 147 IPANARLLCSQPFVEVEKACRSYSHVLEKEVRFNQLSYLLKDPSLNHLIHLVRD 206
 Db 220 VEDTACERSCPVAIRALEACRKYVYVYIKDVLIDGLVPLRLRPGINLRVYOLFRD 279
 QY 207 PRAVFSRERTKGLMIDS-----RIYV-----GHEKLLKKEQPY 244
 Db 280 PRAVHNSRLKSRGGLRESIQVLRTRNGDRFHRVLLAHGVARGPGQSALPAAPADP 339
 QY 245 ----VMOVICOSQLEIYKTIQSLPKALQERYLLVREYEDLARAPVACTSRMYEFVGLFELP 300
 Db 340 FLTGALVEICEMALRDLRFARCAPAWLRRLRYLRFEDLVROPRADRLRLRRSGLRALA 399
 QY 301 HLOTWVHNITRGKMG-DHAFTNARDALNVSOAMRMSLEPEVSRLOKACGDAMNLLGY 359
 Db 400 ALDAFALNMTRGAAAGADREFHLSARDAREAVHAMRELSREQVQVEVACAPAMRLAY 459
 QY 360 RHVRSQOE 367
 Db 460 --PRSGEE 465

RESULT 13

Q98276 PRELIMINARY; PRT; 530 AA.
 AC Q98276; 01-NOV-1998 (Tremblrel, 08, Created)
 DT 01-NOV-1998 (Tremblrel, 08, Last sequence update)
 DE 01-DEC-2001 (Tremblrel, 19, Last annotation update)
 GN N-acetylglucosamine-6-O-sulfotransferase long form.
 OS Mus musculus (mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NCBI_TaxID=10090;
 [1] NCBI_TaxID=10090;
 RP SEQUENCE FROM N.A.
 RC TISSUE=Embryo;
 RX MEDLINE=98380482; PubMed=9712885;
 RA Uchimura K., Muramatsu H., Kadomatsu K., Fan O., Kurosawa N.,
 RA Mitsuoka C., Kannagi R., Habuchi O., Muramatsu T.;
 RT "Molecular cloning and characterization of an N-acetylglucosamine-6-O-
 RT sulfotransferase.";
 RL J. Biol. Chem. 273:22577-22583(1998).
 DR EMBL: AB011452; BAA32138.1; -;
 DR EMBL: AB011452; BAA32139.1; -;
 DR EMBL: AB011451; BAA32137.1; -;
 DR MGI:1891160; Chst2.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Sulfotransferase.
 SQ SEQUENCE 530 AA; 57814 MW; A113E1B735C63EC CRC64;

Query Match

Best Local Similarity 29.4%; Score 598.5; DB 11; Length 530;
 Best Local Similarity 36.8%; Pred. No. 8e-48;
 Matches 134; Conservative 64; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERNHVVLSSMRSGSFFVQLFGQHPDVFYIMEPAHVMTFKOSTAMMLHMAVRDLIRA 100
 Db 163 KROLVYVFTWTRSGSFFGLFNOHPDVFYIMEPAHVMTFKOSTAMMLHMAVRDLIRA 222
 QY 101 VFLCDMSVFDAYMEPGRRS-----SLEFQENSRAICAPACDIIPDEIIPAHRCLL- 155
 Db 223 LYRCDPSVLTALYRPGDPAARAPDANLTLTAALFRWNTNKVICSPLCPAPARAREVGL 281
 QY 156 -CSQPFVEVEKACRSYSHVLEKEVRFNQLSYLLKDPSLNHLIHLVRDPAVRSR 214
 Db 282 KCPQRLARREECKRKYVYVYIKDVLIDGLVPLRLRPGINLRVYOLFRD 341
 QY 215 ERTKGLMIDSRIVM-----GHEKLLKKEQPY---PY---VMOVICOS 252
 Db 342 IRRHGLIRSLQVRSRDPRAHMPLEAAGHKLGAKEGKMGADYHALGMEVYICNS 401
 QY 253 QLEIYKTIQSLPKALQERYLLVREYEDLARAPVACTSRMYEFVGLFELPQLTWHNTTRG 312
 Db 402 MARTLQALPPLMGHLYVREYEDLVGPVTLRKYVDFGLVLSPEMEQPALNMTSG 461
 QY 313 KGMGDHAFHTNARDALNVSOAMRMSLEPEVSRLOKACGDAMNLLGYRHVRSQOEORNL 372
 Db 462 SGSSKPEFVVSARNATQANAMRTALTFOQIKOYEEFCYPMALGTERVNSPEEKDLS 521
 QY 373 LDLL 376
 Db 522 KTL 525

RESULT 14

Q9UED5 PRELIMINARY; PRT; 483 AA.
 AC Q9UED5; 01-MAR-2003 (Tremblrel, 13, Created)
 DT 01-MAR-2003 (Tremblrel, 23, Last sequence update)
 DE 01-MAR-2003 (Tremblrel, 23, Last annotation update)
 GN N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST).
 QY 460 --PRSGEE 465

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:41:19 ; Search time 16.4574 Seconds
(without alignments)
992.382 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLPRKKMKLLFLVSQMAIL.....EQRNLLDLSTWVPEQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 4231058 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/1/1aa/5A-COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B-COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A-COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B-COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCTUS-COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2038	100.0	386	3 US-09-045-284A-2	Sequence 2, Appl1
2	2038	100.0	386	4 US-09-190-911-1	Sequence 1, Appl1
3	1936	95.0	386	4 US-09-786-240-11	Sequence 11, Appl1
4	598.5	29.4	483	3 US-09-263-023-2	Sequence 2, Appl1
5	598.5	29.4	483	3 US-09-471-867-2	Sequence 2, Appl1
6	587.5	28.8	484	3 US-09-263-023-4	Sequence 4, Appl1
7	587.5	28.8	484	4 US-09-471-867-4	Sequence 4, Appl1
8	549	26.9	411	4 US-09-015-188-2	Sequence 2, Appl1
9	500.5	24.6	438	2 US-08-655-878-2	Sequence 2, Appl1
10	482	23.7	479	2 US-08-899-514-2	Sequence 2, Appl1
11	100.5	4.9	668	4 US-09-811-469-6	Sequence 27, Appl1
12	94.5	4.6	848	4 US-09-540-824-27	Sequence 27, Appl1
13	92	4.5	359	3 US-09-150-133-9	Sequence 11, Appl1
14	92	4.5	359	3 US-09-150-141-11	Sequence 11, Appl1
15	92	4.5	359	3 US-09-374-493-11	Sequence 11, Appl1
16	92	4.5	359	3 US-09-374-824-11	Sequence 11, Appl1
17	92	4.5	359	3 US-09-374-492-11	Sequence 11, Appl1
18	87.5	4.3	380	3 US-09-150-133-9	Sequence 9, Appl1
19	87.5	4.3	380	3 US-09-150-141-9	Sequence 9, Appl1
20	87.5	4.3	380	3 US-09-374-493-9	Sequence 9, Appl1
21	87.5	4.3	380	3 US-09-374-824-9	Sequence 9, Appl1
22	87.5	4.3	380	3 US-09-374-492-9	Sequence 9, Appl1
23	86.5	4.2	831	1 US-08-073-384C-5	Sequence 5, Appl1
24	86.5	4.2	831	1 US-08-254-359A-C-5	Sequence 5, Appl1
25	86.5	4.2	831	1 US-08-483-043-5	Sequence 5, Appl1
26	86.5	4.2	831	1 US-08-483-238-5	Sequence 5, Appl1
27	86.5	4.2	831	2 US-08-471-066B-5	Sequence 5, Appl1

28	86.5	4.2	831	2 US-08-484-956-5	Sequence 5, Appl1
29	86.5	4.2	831	2 US-08-757-653-5	Sequence 5, Appl1
30	86.5	4.2	831	2 US-08-599-491-5	Sequence 5, Appl1
31	86.5	4.2	831	2 US-08-756-386-5	Sequence 5, Appl1
32	86.5	4.2	831	2 US-08-823-516-5	Sequence 5, Appl1
33	86.5	4.2	831	3 US-08-682-853A-5	Sequence 5, Appl1
34	86.5	4.2	831	3 US-08-759-038-5	Sequence 5, Appl1
35	86.5	4.2	831	3 US-08-758-314-5	Sequence 5, Appl1
36	86.5	4.2	831	4 US-09-350-309-5	Sequence 5, Appl1
37	86.5	4.2	831	4 US-08-520-946-5	Sequence 5, Appl1
38	86.5	4.2	831	4 US-09-684-938-5	Sequence 5, Appl1
39	86.5	4.2	831	4 US-09-308-825A-5	Sequence 5, Appl1
40	84.5	4.1	2549	4 US-08-265-967C-1	Sequence 1, Appl1
41	84.5	4.1	2549	4 US-08-305-790B-2	Sequence 2, Appl1
42	83	4.1	195	4 US-09-252-991A-30447	Sequence 30847, A
43	83	4.1	222	2 US-08-485-721-9	Sequence 9, Appl1
44	83	4.1	222	2 US-08-392-935-9	Sequence 9, Appl1
45	83	4.1	222	5 PCT-US93-08326-9	Sequence 9, Appl1

ALIGNMENTS

RESULT 1					
US-09-045-284A-2					
: Sequence 2, Application US/09045284A					
: Patent No. 6265192					
: GENERAL INFORMATION:					
: APPLICANT: Blstrup, Annette					
: APPLICANT: Rosen, Steven D.					
: APPLICANT: Hemmerich, Stefan					
: TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3					
: FILE REFERENCE: 6510-107US1					
: CURRENT APPLICATION NUMBER: US/09/045,284A					
: CURRENT FILING DATE: 1998-03-20					
: NUMBER OF SEQ ID NOS: 9					
: SOFTWARE: FastSeq for Windows Version 4.0					
: SEQ ID NO 2					
: LENGTH: 386					
: TYPE: PRT					
: ORGANISM: Homo sapiens					
US-09-045-284A-2					
Query Match 100.0%; Score 2038; DB 3; Length 386;					
Best Local Similarity 100.0%; Pred. No. 2, 3e-212;					
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MLPRKKMKLLFLVSQMAILALFFHMTSHNITSLSKRAQPERHVLVLSRRSGSFTVGQ	60		
DB	1	MLPRKKMKLLFLVSQMAILALFFHMTSHNITSLSKRAQPERHVLVLSRRSGSFTVGQ	60		
QY	61	LEGGHPVEVFLMEPRAMVMMTFKOSTAMMLHMAVRDILRAVFLCDMSVFPAYMGRPGRQ	120		
DB	61	LEGGHPVEVFLMEPRAMVMMTFKOSTAMMLHMAVRDILRAVFLCDMSVFPAYMGRPGRQ	120		
QY	121	SSLEQWENSRLCASAPACDIIPODEIIPRACHRLCSQPFVEVEKACRSYSHVLEVR	180		
DB	121	SSLEQWENSRLCASAPACDIIPODEIIPRACHRLCSQPFVEVEKACRSYSHVLEVR	180		
QY	181	FFNIQSLYPLKDPSSLNHLVHLYDRPRAVFRSEPRKGDIMDSRIYMGHEOKLKED	240		
DB	181	FFNIQSLYPLKDPSSLNHLVHLYDRPRAVFRSEPRKGDIMDSRIYMGHEOKLKED	240		
QY	241	OPYVMOVITCOSOLEIKTTIOSLPKAOERLLVRYEDLARAPVAOTSRMYEFGLEFLP	300		
DB	241	OPYVMOVITCOSOLEIKTTIOSLPKAOERLLVRYEDLARAPVAOTSRMYEFGLEFLP	300		
QY	301	HLQTVWNITRGKMGDHAFTNARDALNVSQARWLSLPYEKVSRLQKAGDANMLLGYR	360		
DB	301	HLQTVWNITRGKMGDHAFTNARDALNVSQARWLSLPYEKVSRLQKAGDANMLLGYR	360		
QY	361	HVNSQEQRNLLDLSTWVPEQIH	386		
DB	361	HVNSQEQRNLLDLSTWVPEQIH	386		

Db 361 HVRSDEQRNLLDLSTWVPEQIH 386

RESULT 2
US-09-190-911-1
; Sequence 1, Application US/09190911
; Patent No. 635365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSTL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 386
; TYPE: PRT
; ORGANISM: H. sapiens
US-09-190-911-1

Query Match 100.0%; Score 2038; DB 4; Length 386;
Best Local Similarity 100.0%; Pred. No. 2,3e-212;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLPPKMKLLFLVSOMATLALFPHYSHNISLSKKAQPERMHVLYSSMRSGSSEFGQ 60
DB 1 MLPPKMKLLFLVSOMATLALFPHYSHNISLSKKAQPERMHVLYSSMRSGSSEFGQ 60
QY 61 LFGQHPDYVYLMEPAMHVMTEFKOSTAWMLHNAVRLJRAVFLCDMSVFDAYMEGPQRQ 120
DB 61 LFGQHPDYVYLMEPAMHVMTEFKOSTAWMLHNAVRLJRAVFLCDMSVFDAYMEGPQRQ 120
QY 121 SSLFOWENSRALCSAPACDIIPQDEIIPRAHRLCSQOPFEVEKACRSYSHVYLKEVR 180
DB 121 SSLFOWENSRALCSAPACDIIPQDEIIPRAHRLCSQOPFEVEKACRSYSHVYLKEVR 180
QY 181 FENLOSIVPLKDPSPNLHIVLVDPRAVFRSRETKDMLIDSRIYVGOHEOKLKED 240
DB 181 FENLOSIVPLKDPSPNLHIVLVDPRAVFRSRETKDMLIDSRIYVGOHEOKLKED 240
QY 241 QPYVWQVICSQOLEIYKTIQSLPKALOERYLLVREYEDLARAFAVQTSMYEFGLEFLP 300
DB 241 QPYVWQVICSQOLEIYKTIQSLPKALOERYLLVREYEDLARAFAVQTSMYEFGLEFLP 300
QY 301 HIQTWVHNITRGKMGMDHAFHTNARDALNVSOAMRWSLPEYKVSRLQKACGDMNLLGYR 360
DB 301 HIQTWVHNITRGKMGMDHAFHTNARDALNVSOAMRWSLPEYKVSRLQKACGDMNLLGYR 360
QY 361 HVRSDEQRNLLDLSTWVPEQIH 386
DB 361 HVRSDEQRNLLDLSTWVPEQIH 386

RESULT 3
US-09-786-240-11
; Sequence 11, Application US/09786240
; Patent No. 6558935
; GENERAL INFORMATION:
; APPLICANT: INCYTE PHARMACEUTICALS, INC.
; APPLICANT: TANG, Y. TOM
; APPLICANT: CORLEY, Neil C.
; APPLICANT: GUEGLER, Karl J.
; APPLICANT: BAUGHN, Mariah R.
; APPLICANT: LAL, Preeti
; APPLICANT: YUE, Henry
; APPLICANT: HILTMAN, Jennifer L.
; APPLICANT: AZIMZAI, Valda

; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592 PCT
; CURRENT APPLICATION NUMBER: US/09/786,240
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/150,657; unaassigned: 09/186,779; unaassigned: 60/133,6
; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1
US-09-786-240-11

Query Match 95.0%; Score 1936; DB 4; Length 386;
Best Local Similarity 95.6%; Pred. No. 2,6e-201;
Matches 369; Conservative 3; Mismatches 14; Indels 0; Gaps 0;

QY 1 MLPPKMKLLFLVSOMATLALFPHYSHNISLSKKAQPERMHVLYSSMRSGSSEFGQ 60
DB 1 MLPPKMKLLFLVSOMATLALFPHYSHNISLSKKAQPERMHVLYSSMRSGSSEFGQ 60
QY 61 LFGQHPDYVYLMEPAMHVMTEFKOSTAWMLHNAVRLJRAVFLCDMSVFDAYMEGPQRQ 120
DB 61 LFGQHPDYVYLMEPAMHVMTEFKOSTAWMLHNAVRLJRAVFLCDMSVFDAYMEGPQRQ 120
QY 121 SSLFOWENSRALCSAPACDIIPQDEIIPRAHRLCSQOPFEVEKACRSYSHVYLKEVR 180
DB 121 SSLFOWENSRALCSAPACDIIPQDEIIPRAHRLCSQOPFEVEKACRSYSHVYLKEVR 180
QY 181 FENLOSIVPLKDPSPNLHIVLVDPRAVFRSRETKDMLIDSRIYVGOHEOKLKED 240
DB 181 FENLOSIVPLKDPSPNLHIVLVDPRAVFRSRETKDMLIDSRIYVGOHEOKLKED 240
QY 241 QPYVWQVICSQOLEIYKTIQSLPKALOERYLLVREYEDLARAFAVQTSMYEFGLEFLP 300
DB 241 QPYVWQVICSQOLEIYKTIQSLPKALOERYLLVREYEDLARAFAVQTSMYEFGLEFLP 300
QY 301 HIQTWVHNITRGKMGMDHAFHTNARDALNVSOAMRWSLPEYKVSRLQKACGDMNLLGYR 360
DB 301 HIQTWVHNITRGKMGMDHAFHTNARDALNVSOAMRWSLPEYKVSRLQKACGDMNLLGYR 360
QY 361 HVRSDEQRNLLDLSTWVPEQIH 386
DB 361 HVRSDEQRNLLDLSTWVPEQIH 386

RESULT 4
US-09-263-023-2
; Sequence 2, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA41.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; CURRENT FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-54007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2


```

Db      296  IRRRHGLIRESIQVYRSRDPRAHRMPFLEAGHKIGAKKEGVCBPADYHALGAMEVICNS 355
OY      253  QLEIKYKTSUPKALQERLYVRYEDLARAQAQTSRMVEFGLEFPLHLOTWVHNITRG 312
Db      356  MAKTIQIALQPPDMIQGHILVYREDLVGDPVKTLRRYIDFVGLVSPMEQFALNMVSG 415
OY      313  KMGDHAFTNARDALNVSOAWRMSLPEKYVSRLOKACGDAMNLLGRRHVRSEQORNL 372
Db      416  SSSSSKPFVYSARNMTQAANMARTALTTPQIQKOVEEFCQPAVAYIGYEVNPSPEVKDLS 475
OY      373  IDLL 376
Db      476  KTL 479

```

RESULT 7

```

-09-471-867-4
Sequence 4, Application US/09471867
Patent No. 6455289

```

```

GENERAL INFORMATION:
APPLICANT: Uchimura, Kenji
APPLICANT: Muramatsu, Hideki
APPLICANT: Kadomatsu, Kenji
APPLICANT: Kamegaki, Reiji
APPLICANT: Habuchi, Osami
APPLICANT: Muramatsu, Takashi
TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
FILE REFERENCE: TOYAMA1.001AUS
CURRENT APPLICATION NUMBER: US/09/411,867
CURRENT FILING DATE: 1999-12-23
EARLIER APPLICATION NUMBER: US 09/263,023
EARLIER FILING DATE: 1999-03-05
EARLIER APPLICATION NUMBER: JP 10-54007
EARLIER FILING DATE: 1998-03-05
EARLIER APPLICATION NUMBER: JP 10-177844
EARLIER FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 4
LENGTH: 484
TYPE: PRF
ORGANISM: Homo sapiens
US-09-471-867-4

```

```

Query Match      28.8%; Score 587.5; DB 4; Length 484;
Best Local Similarity 36.0%; Pred. No. 5,6e-55;
Matches 131; Conservative 67; Mismatches 137; Indels 29; Gaps 6;

```

```

OY      41  EEMHVLVLSRSGSFGQLFGQHPDVYFLMEPAHVMYMTKOSTAMMLHNAVRLIRA 100
Db      117  KHNMYVFTTWRSFGSFGELFNONPEVFLPEVHWOKLPGDAVSLQGAARDMLSA 176
OY      101  VFLCDMSVFDAYMEGPRROS----SLFOWENSRLCSAPACDIIPODEIIPRACHRL- 155
Db      177  LYRCLSLVQLYSPAGSGGRNLTIGIFGAATNKVYCSSPLCPAY-KREVLVLVDRCK 235
OY      156  -CSQGFVEVYKACRSYSHVLEKEVFNQSLYPLKDPSSLNHLIHLVLRPRAVRSR 214
Db      236  KCPQRLARFECECKRYTLVIGVRFDAVLAFLRLPDALDLVYHLVRRPRAVRSR 295
OY      215  ERTKDDMLDSRTVM-----GOHEOKLKE---DQPIY--VMQYICOS 252
Db      296  IRRRHGLIRESIQVYRSRDPRAHRMPFLEAGHKIGAKKEGVCBPADYHALGAMEVICNS 355
OY      253  QLEIKYKTSUPKALQERLYVRYEDLARAQAQTSRMVEFGLEFPLHLOTWVHNITRG 312
Db      356  MAKTIQIALQPPDMIQGHILVYREDLVGDPVKTLRRYIDFVGLVSPMEQFALNMVSG 415
OY      313  KMGDHAFTNARDALNVSOAWRMSLPEKYVSRLOKACGDAMNLLGRRHVRSEQORNL 372
Db      416  SSSSSKPFVYSARNMTQAANMARTALTTPQIQKOVEEFCQPAVAYIGYEVNPSPEVKDLS 475

```

```

OY      373  IDLL 376
Db      476  KTL 479

```

RESULT 8

```

US-09-015-188-2
Sequence 2, Application US/09015188C
Patent No. 6393358

```

```

GENERAL INFORMATION:
APPLICANT: Williams, Kevin J
APPLICANT: Tabas, Ira
TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
FILE REFERENCE: JEPF-0231
CURRENT APPLICATION NUMBER: US/09/015,188C
CURRENT FILING DATE: 1998-01-29
NUMBER OF SEQ ID NOS: 17
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 2
LENGTH: 411
TYPE: PRF
ORGANISM: Homo sapiens
US-09-015-188-2

```

```

Query Match      26.9%; Score 549; DB 4; Length 411;
Best Local Similarity 34.3%; Pred. No. 6,5e-51;
Matches 139; Conservative 64; Mismatches 160; Indels 42; Gaps 10;

```

```

OY      6  KKLLFLVS---QMAITLALFPMTSHNISLSMKAQPERM-----HYLV 47
Db      6  KAVLLLAALASIAIQYTAIRTFYAKSFHPCGIAEGLAEERCEESPTAYNLSRKTILI 65
OY      48  ISSRSGSFGQLFGQHPDVYFLMEPAHVMYMTKOSTAMMLHNAVRLIRA 99
Db      66  LATTSSGSSFGQLFNQHLVDVYFLPEPLXHYQNTILPFTQKSRADRYMAGASRDLLR 125
OY      100  AVFLCDMSVFDAYMEGPRROS--LFOWENSRLCSAPACDIIPODEIIPRACHRLC 156
Db      126  SLYDCDLVLENYIKPPVNHHTDRIFRGASRVLCSPVCDPPGADLVLEEGCVKRC 185
OY      157  SQGFVEVYKACRSYSHVLEKEVFNQSLYPLKDPSSLNHLIHLVLRPRAVRSR 216
Db      186  GILNLTVAEACRESHVAIKTVRVENDLALVEDPRLNKVYQLVRDPGGLIASNSE 245
OY      217  TKGDLMDISRIYMGQHEOKLKKEDQPYVY---MOVICOSLEIKYKTSUPKALQERYL 272
Db      246  TFRDYYRLRLMYGGR-----KPYNLDVQLTTCVEDFNSYSTGLMRPMLKGYM 298
OY      273  LVRYEDLARAQAQTSRMVEFGLEFPLHLOTWVHNITRG-KMGDHAFTNARDALNV 331
Db      299  LVRYEDLARNPKKTEETIYIGFIPDLSHVAVMIQNNRGPDTLCKHXYGT-VRSAAATA 357
OY      332  QAWRMSLPEKYVSRLOKACGDAMNLLGRRHVRSEQORNLIDLL 376
Db      358  EKMRFRLSTDYVAFQANACQOVLADQGYKIAASEBELANPSVSLV 402

```

RESULT 9

```

US-08-655-878-2
Sequence 2, Application US/08655878
Patent No. 5827713

```

```

GENERAL INFORMATION:
APPLICANT: FUKUTA, MASAKAZU
APPLICANT: HABUCHI, OSAMI
TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESS:
STREET:
CITY:
STATE:

```

```

: COUNTRY:
: ZIP:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/655,878
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME:
: REGISTRATION NUMBER:
: REFERENCE/DOCKET NUMBER:
: TELECOMMUNICATION INFORMATION:
: TELEPHONE:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 458
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-655-878-2

```

Query Match 24.6%; Score 500.5; DB 2; Length 458;

Best Local Similarity 33.2%; Pred. No. 1.4e-45;

Matches 120; Conservative 75; Mismatches 133; Indels 33; Gaps 13;

```

QY 34 LSKADPERHNVIVLSRSGSFVGLFGQHPDVFLMPAHV--WTFKOSTAMLH 91
DB 104 LGIAAPERPRHVLMTATRTGSSVFGEFNQGNIFLFEPLMHIERTVFEPEGANAVG 163
QY 92 MAV--RDLIRAVFLCDMSVFDAYMEPRRQ--SLPQWENSRLCSAPACDIIPQ-DEI 146
DB 164 SALVYRNVLOQLLDCDIYIESFTSPAPBEHLRALFRSSSHSLCEPVC--TPSLKV 221
QY 147 IPRAHCR-LICSQQPEVEVEKACRSYSHVLAKEVRFNQLSLYPLKPSLNLHIVLVR 205
DB 222 FEYHYCHNRRCGPLNITLAEACRRKQHMALKTVRIRQLFLOPLADPRDLRIQLVLR 281
QY 206 DPRVAFRSRERTGDLMDIRIVMGHEQK-----KEQPIYVMOVICS-QLE 255
DB 282 DPRVAVLSR-----MV--AFSGKYESMKMAAEGLAEPLQDE-VQRLGNCEISRLS 330
QY 256 IYKTIOSLPKALOERYLLVREEDLARAPVAQTSRMVEFVLEFLPHLQTVVHNITRGKM 315
DB 331 AELGLNQ-PRMLGRIVLVREEDYARAPLKALEMIFAGIHPPQVEEIRANTQAP-Q 388
QY 316 GDHAFHTNADALNVSQAWMSLPEYKVSRLQKACGDAMNLGGRHVSQEQQRNLLDL 375
DB 389 DSGIYSTOKNSSEOFKMFPSIPFKLAQVVOACBPAMRLFGYKLASSAQELTNRLSL 448
QY 376 L 376
DB 449 L 449

```

RESULT 10

US-08-899-514-2

; Sequence 2, Application us/08899514

; Patent No. 5910581

; GENERAL INFORMATION:

; APPLICANT: HABUCHI, OSAMI

; APPLICANT: FUKUTA, MASAKAZU

; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN

; TITLE OF INVENTION: SUBOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING

; NUMBER OF SEQUENCES: 9

```

: CORRESPONDENCE ADDRESS:
: ADDRESSEE: KNOBE, MARTENS, OLSON & BEAR, LLP
: STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
: CITY: NEWPORT BEACH
: STATE: CALIFORNIA
: COUNTRY: US
: ZIP: 92660
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/899,514
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: DANIEL E. ALTMAN
: REGISTRATION NUMBER: 34,115
: REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 714 760 0404
: TELEFAX: 714 760 9502
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 479
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-899-514-2

```

Query Match 23.7%; Score 482; DB 2; Length 479;

Best Local Similarity 32.4%; Pred. No. 1.5e-43;

Matches 113; Conservative 71; Mismatches 141; Indels 24; Gaps 10;

```

QY 42 RMHVLVLSRSGSFVGLFGQHPDVFLMPAHVWMT--PROGAMMLHNAV--RDL 97
DB 132 RRHVLMTATRTGSSVFGEFNQGNIFLFEPLMHIERTVFEPEGANAGSALVYRDV 191
QY 98 IRAVFLCDMSVFDAYMEPRRQSSLPQWE--NSRALCSAPACDIIPQDEIIPRAHCR-L 154
DB 192 LKQFLDCDIYLEHFTPLDEHDTQFMFRGSSRLCEDPVCIPFK-KVEEYHCKNR 250
QY 155 LCSQQPEVEVEKACRSYSHVLAKEVRFNQLSLYPLKPSLNLHIVLVRDPRVFRSR 214
DB 251 RCGPLNVLTAAEACRRKREHMLKAVIRQLFLOPLADPRDLRIQLVDRPRAVLASR 310
QY 215 -----ERTGDLMDIRIVMGHEQKLEKEDQPIYVMOVICS-QLEIYKTIOSLPKAL 267
DB 311 MVAEPAGKYTKWKMLDDEGQDLNEEVOR-----LRNCESIRLSAELGLNQ-PAWL 362
QY 268 QERYLVREEDLARAPVAQTSRMVEFVLEFLPHLQTVVHNITRGKMGHAFHTNARDA 327
DB 363 RGRMLVRYEDVARGPLQKAREMTPEAGIPLTPOVEDMICKNQNAADSGISTQ-KNS 421
QY 328 LNVSQAWMSLPEYKVSRLQKACGDAMNLGGRHVSQEQQRNLLDL 376
DB 422 SEQEFKWFSPMFKLAQVVOACBPAMRLFGYKLARDAALTNRSVSL 470

```

RESULT 11

US-09-811-469-6

; Sequence 6, Application us/09811469

; Patent No. 6531809

; GENERAL INFORMATION:

; APPLICANT: YAN, Chunhua et al.

; TITLE OF INVENTION: ISOLATED HUMAN PHOSPHATASE PROTEINS,

; TITLE OF INVENTION: NUCLEIC ACID MOLECULES ENCODING HUMAN PHOSPHATASE PROTEINS,

; FILE REFERENCE: C1001171

;; CURRENT APPLICATION NUMBER: US/09/811,469
 ;; CURRENT FILING DATE: 2001-03-20
 ;; NUMBER OF SEQ ID NOS: 11
 ;; SOFTWARE: FastSeq for Windows Version 4.0
 ;; SEQ ID NO 6
 ;; LENGTH: 668
 ;; TYPE: prt
 ;; ORGANISM: Human
 US-09-811-469-6

Query Match 4.9%; Score 100.5; DB 4; Length 668;
 Best Local Similarity 22.7%; Pred. No. 0.056;
 Matches 57; Conservative 35; Mismatches 94; Indels 65; Gaps 12;

QY 156 CSQPEVEVERKASYSHYVLEKEVREFNLSLYPLKDPSLMLHIYLVDRPRAVRSRE 215
 DB 356 CCKKAEIECEAEANMNVLLLEE--NASDCCILSS-----LVQKMDPHC-----RT 402
 DB 216 RTKGDLMTDSRIYMG-----QHOKIKKEDQPYVYVQVICSQSLYEIKTISLPAKL 267
 DB 403 RIGFOSLQKEVWGGHCFDLRCNHLRQDKKEVPFELLFDC-----VWQLVHOPPAF 457
 QY 268 Q--ERYLVREYEDLARAPVQTSRMVEFVGLFELPHLQTVNHNTERGK--GMDHAFHTN 323
 DB 458 EETETLYLVLSDSL-----YIPISFEPFNSPHQKDTNMGREGQDPTQ 499
 QY 324 ABDALNVSQARMSLPEY-KVSRLOKACGDAMNL-----GYRVRSEQEQRNILL 373
 DB 500 SK-PNLTLVWDMVSQVFEKQATLLK-----NPLYVEKPKLDKGRKMRKFOROLSL 552
 QY 374 DLTSTWVPEQ 384
 DB 553 PLTOSKSSPKR 563

RESULT 12
 US-09-540-824-27
 ; Sequence 27, Application US/09540824
 ; Patent No. 6383753
 ; GENERAL INFORMATION:
 ; APPLICANT: Thiele, Dennis
 ; APPLICANT: Liu, Phillip
 ; TITLE OF INVENTION: No. 6383753el Yeast and Mammalian Regulators of Cell Proliferat
 ; FILE REFERENCE: UW-04266
 ; CURRENT APPLICATION NUMBER: US/09/540,824
 ; CURRENT FILING DATE: 2000-03-31
 ; NUMBER OF SEQ ID NOS: 28
 ; SOFTWARE: PatentIn version 3.0
 ; SEQ ID NO 27
 ; LENGTH: 848
 ; TYPE: prt
 ; ORGANISM: Caenorhabditis elegans
 US-09-540-824-27

Query Match 4.6%; Score 94.5; DB 4; Length 848;
 Best Local Similarity 21.2%; Pred. No. 0.36;
 Matches 74; Conservative 46; Mismatches 130; Indels 99; Gaps 15;

QY 5 KMKLLFLVSOMATLALFFHMYSHNISLSMKAQ---PERMHVVLSSMRGSSSFVQOL 61
 DB 426 KRLKMLFL-----CLRWDMDSKBELTAESYKKNLADQETMALMKPDIQGVGRVRL 481
 QY 62 FGOH-----PDVEYLMPEAMHWMTFKQSTAMLMHNAVRLIRAVFLCDMSVFD 110
 DB 482 IQHWRKRODKQKSSPVSFGLISAIRLVSGLEFPVADSV--HPVV--VPALFLATEALCS 536
 QY 111 A-----YMEGPRROSSLPQWENSRALCSAP-----ACDIIPD 144
 DB 537 AACANLMLAKQIOLANALIVEYSESKRYPPELVAFARSALLATYEKSEKPAITNGFP-- 594
 QY 145 EIIPIAHCRLLC-----SQQPE-----EVEKACRSYSHVLYKCV 179
 DB 595 --ISKPHMLCFEKKLFLTKNYNMISQKEFDLSSFNRYIDELYLKMTKRTYTGPALOPI 652

QY 180 REFNIQSLYPLKDPSLMLHIYH---LVDRPRAVRSRERTKGDLMID-SRIYMGQHEQ 234
 DB 653 SLTTFINNSP--SDPSLKILHVRALLSTLQHLRVAYSQNETYYSIVFKFPLILSSIQAK 710
 QY 235 KKKKEDQPYVYVQVICS-----QLEIKTISLPAKQERY 271
 DB 711 NLPAEVOE--ELETICASKMAEIGAKRVLVHLSTLVKTERKSMKMLEPRP 757

RESULT 13
 US-09-150-133-11
 ; Sequence 11, Application US/09150133B
 ; Patent No. 6060295
 ; GENERAL INFORMATION:
 ; APPLICANT: The Board of Regents of the University of Oklahoma
 ; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
 ; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
 ; FILE REFERENCE: 5820.504
 ; CURRENT APPLICATION NUMBER: US/09/150,133B
 ; CURRENT FILING DATE: 1998-09-09
 ; NUMBER OF SEQ ID NOS: 45
 ; SOFTWARE: Wordperfect 5.1 (saved in Ascii format)
 ; SEQ ID NO 11
 ; LENGTH: 359
 ; TYPE: prt
 ; ORGANISM: Caenorhabditis elegans
 US-09-150-133-11

Query Match 4.5%; Score 92; DB 3; Length 359;
 Best Local Similarity 19.5%; Pred. No. 0.18;
 Matches 63; Conservative 48; Mismatches 114; Indels 98; Gaps 15;

QY 52 RSGSSFVQGLFGQHPDY---FYLMPEAMHWMTFKQSTAMLMHNAVRLIRAVFLCDMS 107
 DB 95 RSGTFLMRALIDHAPDVRCGETMLPSFLTWQAGMRND-WVNNSGI-----TQE 143
 QY 108 VFDAVMEPPRROSSLPQWENSRALCSAPACIITQ-DEIIRACHRLCSQQPEVVERK 166
 DB 144 VFDDAV-----SAFTTIVAKHSLAPR-----LCNRDP----- 172
 QY 167 ACSRSYHVLEKEVREFNLSLYPLKDPSLMLHIYLVDRPRAVRSRERTK---GDLM 222
 DB 173-----YIALDPLTIR-----RLTP-----NAKFLMITDAVAHVSMIERKVPVAGYWT 216
 QY 223 IDSRIYMGQHEQKLEKEDQPYVYVQVICSQ---LEIYKTISLPAKQERYLVREYD 278
 DB 217 SDEISMFVQWNOELRK-----MTFQCNNAPGQCICKYV-----YER 251
 QY 279 LARAPVQTSRMVEFVGLFELPHLQTVNHNTERGKGMGHAFH-TNARDALNVSQARMS 337
 DB 252 LIQKPAEELIRITNFDLPFSOOMLRHODLIDDEVLDNDEFSASQVKNISINTKALTSWF 311
 QY 338 LPY-EKVSRLQKACGDAMNLGY 359
 DB 312 DCFSEETLAKLDVAPFLGILGT 334

RESULT 14
 US-09-150-141-11
 ; Sequence 11, Application US/09150141B
 ; Patent No. 6071732
 ; GENERAL INFORMATION:
 ; APPLICANT: The Board of Regents of the University of Oklahoma
 ; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, NUCLEIC ACIDS ENCODING
 ; TITLE OF INVENTION: TYROSYLPROTEIN SULFOTRANSFERASES, AND METHODS OF USE THEREOF
 ; FILE REFERENCE: 5820.495
 ; CURRENT APPLICATION NUMBER: US/09/150,141B
 ; CURRENT FILING DATE: 1998-09-09
 ; NUMBER OF SEQ ID NOS: 45
 ; SOFTWARE: Wordperfect 5.1 (saved in Ascii format)
 ; SEQ ID NO 11
 ; LENGTH: 359

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:42:44 ; Search time 20.447 Seconds
(Without alignments)
2241.953 Million cell updates/sec

Title: US-09-645-078-2
Perfect score: 2038
Sequence: 1 MLPRKKMLLFLVSOMAIL.....EQRNLLDLSTWPEQIH 386

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/2/pubppaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubppaa/PCY_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubppaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubppaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubppaa/PCYUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubppaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubppaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubppaa/US09_NEW_PUB.pep:*
12: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubppaa/US10_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubppaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubppaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	2038	100.0	386	9	US-09-816-825-2
2	2038	100.0	386	14	US-10-007-262-1
3	1028	50.4	418	9	US-09-927-602-5
4	1019	50.0	395	9	US-09-927-602-2
5	1013.5	49.7	395	9	US-09-927-602-3
6	1008	49.5	390	9	US-09-927-602-4
7	821	40.3	171	9	US-09-927-602-8
8	598.5	29.4	483	15	US-10-212-933-2
9	587.5	28.8	484	15	US-10-212-933-4
10	519.5	25.5	531	9	US-09-833-790-255
11	519.5	25.5	169	9	US-09-927-602-6
12	511.5	25.1	169	9	US-09-927-602-7
13	336	16.5	169	9	US-09-927-602-9
14	335.5	16.5	179	9	US-09-927-602-10
15	293	14.4	174	9	US-09-927-602-11

16	116	5.7	387	15	US-10-126-279-21	Sequence 21, Appl
17	100.5	4.9	668	11	US-09-811-469-6	Sequence 6, Appl1
18	90	4.4	393	15	US-10-126-279-2	Sequence 2, Appl1
19	89	4.4	320	9	US-09-854-122-20	Sequence 20, Appl1
20	89	4.4	1968	15	US-10-223-070-15	Sequence 15, Appl1
21	88	4.3	2209	10	US-09-902-941-1903	Sequence 1903, Ap
22	88	4.3	2209	10	US-09-849-626-1903	Sequence 1903, Ap
23	88	4.3	2209	15	US-10-017-754-1903	Sequence 1903, Ap
24	86.5	4.2	831	11	US-09-940-244-5	Sequence 5, Appl1
25	86.5	4.2	831	11	US-09-940-925A-5	Sequence 5, Appl1
26	86.5	4.2	831	11	US-09-982-667-3	Sequence 5, Appl1
27	86.5	4.2	831	11	US-09-864-636A-158	Sequence 158, App
28	86.5	4.2	831	11	US-09-941-193A-5	Sequence 5, Appl1
29	86.5	4.2	831	14	US-10-033-297-5	Sequence 5, Appl1
30	86.5	4.2	831	14	US-10-081-806-5	Sequence 5, Appl1
31	86.5	4.2	831	15	US-10-074-328-5	Sequence 5, Appl1
32	86	4.2	824	10	US-09-801-368-312	Sequence 312, App
33	82.5	4.0	388	15	US-10-126-279-6	Sequence 6, Appl1
34	82.5	4.0	777	11	US-09-811-469-2	Sequence 2, Appl1
35	82.5	4.0	1140	9	US-09-950-634-4	Sequence 4, Appl1
36	82.5	4.0	2549	11	US-09-950-634-3	Sequence 3, Appl1
37	81.5	4.0	1463	15	US-10-226-065-3	Sequence 3, Appl1
38	81	4.0	228	11	US-09-813-153-286	Sequence 286, App
39	81	4.0	295	10	US-09-965-529-31	Sequence 31, Appl1
40	81	4.0	295	11	US-09-969-680A-31	Sequence 31, Appl1
41	81	4.0	1100	9	US-09-815-242-11861	Sequence 11861, A
42	79.5	3.9	177	8	US-08-424-5508-53	Sequence 53, Appl
43	79.5	3.9	840	11	US-09-864-636A-574	Sequence 574, App
44	79.5	3.9	1458	15	US-10-203-224-3	Sequence 3, Appl1
45	79	3.9	402	9	US-09-815-242-11350	Sequence 11350, A

ALIGNMENTS

RESULT 1
US-09-816-825-2
Sequence 2, Application US/09816825
Patent No. US20010051370A1
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107CON
CURRENT APPLICATION NUMBER: US/09/816, 825
CURRENT FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/045, 284
PRIOR FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 386
TYPE: PRT
ORGANISM: Homo sapiens
US-09-816-825-2

Query Match	100.0%	Score 2038	DB 9	Length 386
Best Local Similarity	100.0%	Pred. No. 1.9e-202		
Matches 386	Conservative 0	Mismatches 0	Indels 0	Gaps 0
QY	1	MLPRKKMLLFLVSOMAILALFFPMYSHNTSSSMKAQPERMHVLYSSWSSSPVQ	60	
DB	1	MLPRKKMLLFLVSOMAILALFFPMYSHNTSSSMKAQPERMHVLYSSWSSSPVQ	60	
QY	61	LEFGHPDVEFLMEBAHVMWTFKOSTAMMLIMAVRDLIRAVFLCDMSVFAYMEGRQ	120	
DB	61	LEFGHPDVEFLMEBAHVMWTFKOSTAMMLIMAVRDLIRAVFLCDMSVFAYMEGRQ	120	
QY	121	SSLFQWENSRLCAAPACDIIPDEIIPRAHCRLLCSQOPREVEKCRSISHYLKEVR	180	
DB	121	SSLFQWENSRLCAAPACDIIPDEIIPRAHCRLLCSQOPREVEKCRSISHYLKEVR	180	

```

QY 181 FENLOSIVLLKDPISLNIHVLVRDPAVRSRETKGDLMDISRTVNGOHEQKLKED 240
DB 181 FENLOSIVLLKDPISLNIHVLVRDPAVRSRETKGDLMDISRTVNGOHEQKLKED 240
QY 241 OPIYVWQVICSQOLEIKTQISLPKALORRYLLVREDLARAFAVQTSRYEFGLEFLP 300
DB 241 OPIYVWQVICSQOLEIKTQISLPKALORRYLLVREDLARAFAVQTSRYEFGLEFLP 300
QY 301 HLOTWVHNITRGKMGDAFHTNARDALNVSOAMWSLPEYKVSRLQKACGDAMNLGGR 360
DB 301 HLOTWVHNITRGKMGDAFHTNARDALNVSOAMWSLPEYKVSRLQKACGDAMNLGGR 360
QY 361 HVRSEOEQRNLLDLSTVTPVPOIH 386
DB 361 HVRSEOEQRNLLDLSTVTPVPOIH 386

```

RESULT 2

```

S-10-007-262-1
Sequence 1, Application US/10007262
Publication No. US20020164748A1
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Tangemann, Kirsten
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107CIP
CURRENT APPLICATION NUMBER: US/10/007,262
CURRENT FILING DATE: 2001-11-08
PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 1
LENGTH: 386
TYPE: PRF
ORGANISM: H. sapiens
US-10-007-262-1

```

```

Query Match 100.0%; Score 2038; DB 14; Length 386;
Best Local Similarity 100.0%; Pred. No. 1.9e-202;
Matches 386; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 1 MLPPKMKLLFLVSOAMAILALFFHMTSHNISLSKKAOPERMHVLYLSSWRSQSSFEVQ 60
b 1 MLPPKMKLLFLVSOAMAILALFFHMTSHNISLSKKAOPERMHVLYLSSWRSQSSFEVQ 60
QY 61 LFGQHPDVYLLMEPAHVMWTFKOSTAMMLHMAVRODLIRAVFLCDMSVFDAYMEPGRRO 120
DB 61 LFGQHPDVYLLMEPAHVMWTFKOSTAMMLHMAVRODLIRAVFLCDMSVFDAYMEPGRRO 120
QY 121 SSLFQWENSRLCASAPACDIIPODEIIPRAHCRLLCSQPFVEVEKACRSYSHVYLKEVR 180
DB 121 SSLFQWENSRLCASAPACDIIPODEIIPRAHCRLLCSQPFVEVEKACRSYSHVYLKEVR 180
QY 181 FENLOSIVLLKDPISLNIHVLVRDPAVRSRETKGDLMDISRTVNGOHEQKLKED 240
DB 181 FENLOSIVLLKDPISLNIHVLVRDPAVRSRETKGDLMDISRTVNGOHEQKLKED 240
QY 241 OPIYVWQVICSQOLEIKTQISLPKALORRYLLVREDLARAFAVQTSRYEFGLEFLP 300
DB 241 OPIYVWQVICSQOLEIKTQISLPKALORRYLLVREDLARAFAVQTSRYEFGLEFLP 300
QY 301 HLOTWVHNITRGKMGDAFHTNARDALNVSOAMWSLPEYKVSRLQKACGDAMNLGGR 360
DB 301 HLOTWVHNITRGKMGDAFHTNARDALNVSOAMWSLPEYKVSRLQKACGDAMNLGGR 360
QY 361 HVRSEOEQRNLLDLSTVTPVPOIH 386
DB 361 HVRSEOEQRNLLDLSTVTPVPOIH 386

```

```

RESULT 3
US-09-927-602-5
Sequence 5, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 5
LENGTH: 418
TYPE: PRF
ORGANISM: Mus musculus
US-09-927-602-5

```

```

Query Match 50.4%; Score 1028; DB 9; Length 418;
Best Local Similarity 56.6%; Pred. No. 8e-98;
Matches 215; Conservative 43; Mismatches 112; Indels 10; Gaps 6;

```

```

QY 1 MLPPK--KMKLLFLVSOAMAILALFFHMTSHNISLSKKAOPERMHVLYLSSWRSQSSFEV 58
DB 25 MRLPRSSVYMLSLWVOTGLVFE--LVSRQVPS-SPAGLGERVHVLYLSSWRSQSSFEV 80
QY 59 GOLFGHPDVYLLMEPAHVMWTFKOSTAMMLHMAVRODLIRAVFLCDMSVFDAYMEPGR 118
DB 81 GOLFGHPDVYLLMEPAHVMWTFKOSTAMMLHMAVRODLIRAVFLCDMSVFDAYL-PMRR 139
QY 119 ROSSLFQWENSRLCASAPACDIIPODEIIPRAHCRLLCSQPFVEVEKACRSYSHVYLKE 178
DB 140 NISDLFQWANSRLCSPYCEAFARONISSEVECKLCTRPGLAGOECSSYSHVYLKE 199
QY 179 VREFNLOSIVLLKDPISLNIHVLVRDPAVRSRETKGDLMDISRTVNGOHEQKLK 238
DB 200 VREFNLOSIVLLKDPISLNIHVLVRDPAVRSRETKGDLMDISRTVNGOHEQKLK 258
QY 239 EDQPIYVWQVICSQOLEIKTQISLPKALORRYLLVREDLARAFAVQTSRYEFGLEFL 298
DB 259 ADPLRLVNVQVCSHVRIAEALHKPPPLDRIYRLVYDLDLPLVIRELYAFTGLCL 318
QY 299 LPHLOTWVHNITRGKMG--DHAFTNARDALNVSOAMWSLPEYKVSRLQKACGDAMNL 356
DB 319 TPOLOTWVHNITRGKMGDAFHTNARDALNVSOAMWSLPEYKVSRLQKACGDAMNL 378
QY 357 LGYRHSOEQRNLLDLSTVTPVPOIH 376
DB 379 LGYRHSOEQRNLLDLSTVTPVPOIH 398

```

```

RESULT 4
US-09-927-602-2
Sequence 2, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 395

```

TYPE: PRT
ORGANISM: Homo Sapien
US-09-927-602-2

Query Match 50.0%; Score 1019; DB 9; Length 395;
Best Local Similarity 54.5%; Pred. No. 6.3e-97;
Matches 216; Conservative 45; Mismatches 107; Indels 28; Gaps 7;

QY 2 LPPKMKLLFLVYSOMAILALFFHMYSHNISLSMKAKPERMHVIVLSSMRSSGSPVQOL 61
DB 14 LLLAQTCLLFLFVSRP-----GPPSPAGCEARVHVIVLSSMRSSGSPVQOL 59
QY 62 FGOHPDVEYLMPEAMHVMTEFKOSTAWMLHMAVVDLIRAVFLCDMSVFDAYMEPGPRROS 121
DB 60 FNOHPDVEYLMPEAMHVMTEFKOSTAWMLHMAVVDLIRAVFLCDMSVFDAYMEPGPRROS 118
QY 122 SLFOWENSRALCSAPACDIITPQDEIIPRAHCRLLCSQOPEVEVEKACRSYSHVYLKEVRF 181
DB 119 DLFOWAVSRALCSAPACDIITPQDEIIPRAHCRLLCSQOPEVEVEKACRSYSHVYLKEVRF 178
QY 182 FNLQSLVPLKDPSSLNHLIYHVRDPAVRSRERTGDMIDSRIVNGOHEOKLKREDQ 241
DB 179 FNLQSLVPLKDPSSLNHLIYHVRDPAVRSRERTGDMIDSRIVNGOHEOKLKREDQ 237
QY 242 PYYWMOVICOSOLEIYK--TIQSLPKALOERYLLVREEDLARAPVQTSRMYEFGLEFL 299
DB 238 GLRVYREVCSHVRIAEAATLKP--PPFLRGYRLVREEDLARAPVQTSRMYEFGLEFL 296
QY 300 PHLOTVHNITRGKMG--DHAFTNARDALNVQAMWSLPEYKVSRLQACGDAMNL 357
DB 297 POLKAMHINITHGSGPGARREAFKTSRNALNVQAMWSLPEYKVSRLQACGDAMNL 356
QY 358 GYRHVRSQEOQRNLLDL-----STWVPEQIH 386
DB 357 GYRHVRSQEOQRNLLDL-----STWVPEQIH 392

RESULT 5

US-09-927-602-3
Sequence 3, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Methods of Treating Macular Corneal
TITLE OF INVENTION: Dystrophy
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 3
LENGTH: 395
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: synthetic construct
NAME/KEY: VARIANT
LOCATION: (1)...(395)
OTHER INFORMATION: Xaa - any amino acid
US-09-927-602-3

Query Match 49.7%; Score 1013.5; DB 9; Length 395;
Best Local Similarity 55.9%; Pred. No. 2.4e-96;
Matches 212; Conservative 45; Mismatches 101; Indels 21; Gaps 6;

QY 2 LPPKMKLLFLVYSOMAILALFFHMYSHNISLSMKAKPERMHVIVLSSMRSSGSPVQOL 61
DB 14 LLLAQTCLLFLFVSRP-----GPPSPAGCEARVHVIVLSSMRSSGSPVQOL 59
QY 62 FGOHPDVEYLMPEAMHVMTEFKOSTAWMLHMAVVDLIRAVFLCDMSVFDAYMEPGPRROS 121

DB 60 FGOHPDVEYLMPEAMHVMTEFKOSTAWMLHMAVVDLIRAVFLCDMSVFDAYMEPGPRROS 118
QY 122 SLFOWENSRALCSAPACDIITPQDEIIPRAHCRLLCSQOPEVEVEKACRSYSHVYLKEVRF 181
DB 119 DLFOWAVSRALCSAPACDIITPQDEIIPRAHCRLLCSQOPEVEVEKACRSYSHVYLKEVRF 178
QY 182 FNLQSLVPLKDPSSLNHLIYHVRDPAVRSRERTGDMIDSRIVNGOHEOKLKREDQ 241
DB 179 FNLQSLVPLKDPSSLNHLIYHVRDPAVRSRERTGDMIDSRIVNGOHEOKLKREDQ 237
QY 242 PYYWMOVICOSOLEIYK--TIQSLPKALOERYLLVREEDLARAPVQTSRMYEFGLEFL 299
DB 238 GLRVYREVCSHVRIAEAATLKP--PPFLRGYRLVREEDLARAPVQTSRMYEFGLEFL 296
QY 300 PHLOTVHNITRGKMG--DHAFTNARDALNVQAMWSLPEYKVSRLQACGDAMNL 357
DB 297 POLKAMHINITHGSGPGARREAFKTSRNALNVQAMWSLPEYKVSRLQACGDAMNL 356
QY 358 GYRHVRSQEOQRNLLDL-----STWVPEQIH 386
DB 357 GYRHVRSQEOQRNLLDL-----STWVPEQIH 375

RESULT 6

US-09-927-602-4
Sequence 4, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
TITLE OF INVENTION: Methods of Treating Macular Corneal
TITLE OF INVENTION: Dystrophy
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 390
TYPE: PRT
ORGANISM: Homo Sapien
US-09-927-602-4

Query Match 49.5%; Score 1008; DB 9; Length 390;
Best Local Similarity 52.3%; Pred. No. 8.6e-96;
Matches 206; Conservative 56; Mismatches 104; Indels 28; Gaps 7;

QY 1 MLEPKMKLLFLVYSOMAILALFFHMYSHNISLSMKAKPERMHVIVLSSMRSSGSPVQOL 60
DB 14 LLLAQTCLLFLFVSRP-----GPPSPAGCEARVHVIVLSSMRSSGSPVQOL 59
QY 61 FGOHPDVEYLMPEAMHVMTEFKOSTAWMLHMAVVDLIRAVFLCDMSVFDAYMEPGPRROS 120
DB 60 FNOHPDVEYLMPEAMHVMTEFKOSTAWMLHMAVVDLIRAVFLCDMSVFDAYMEPGPRROS 118
QY 121 SLFOWENSRALCSAPACDIITPQDEIIPRAHCRLLCSQOPEVEVEKACRSYSHVYLKEVRF 180
DB 119 DLFOWAVSRALCSAPACDIITPQDEIIPRAHCRLLCSQOPEVEVEKACRSYSHVYLKEVRF 178
QY 181 FNLQSLVPLKDPSSLNHLIYHVRDPAVRSRERTGDMIDSRIVNGOHEOKLKREDQ 240
DB 179 FNLQSLVPLKDPSSLNHLIYHVRDPAVRSRERTGDMIDSRIVNGOHEOKLKREDQ 237
QY 241 QPYWMOVICOSOLEIYK--TIQSLPKALOERYLLVREEDLARAPVQTSRMYEFGLEFL 298
DB 238 GLRVYREVCSHVRIAEAATLKP--PPFLRGYRLVREEDLARAPVQTSRMYEFGLEFL 296
QY 300 PHLOTVHNITRGKMG--DHAFTNARDALNVQAMWSLPEYKVSRLQACGDAMNL 356
DB 297 POLKAMHINITHGSGPGARREAFKTSRNALNVQAMWSLPEYKVSRLQACGDAMNL 356

QY 357 LGYRHVRSQEQBNLLDL-----STWYPE 383
 Db 357 LGYRPVYSADQORDLTLVLPRGPDHFSWASPD 390

RESULT 7
 US-09-927-602-8
 ; Sequence 8, Application US/09927602
 ; Patent No. US20020061562A1
 ; GENERAL INFORMATION:

APPLICANT: Fukuda, Michiko N.

APPLICANT: Akama, Tomoya O.

TITLE OF INVENTION: Methods of Treating Macular Corneal

FILE REFERENCE: P-LJ 4852

CURRENT APPLICATION NUMBER: US/09/927,602

CURRENT FILING DATE: 2001-08-09

PRIOR APPLICATION NUMBER: US 09/638,211

PRIOR FILING DATE: 2000-08-11

NUMBER OF SEQ ID NOS: 38

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 8

LENGTH: 171

TYPE: PRT

ORGANISM: Homo Sapien

US-09-927-602-8

Query Match 40.3%; Score 821; DB 9; Length 171;
 Best Local Similarity 60.0%; Pred. No. 6, 2e-77;
 Matches 171; Conservative 0; Mismatches 0; Indels 114; Gaps 2;

QY 32 SSLSMAQPEBKHVYLVSSRSGSSFFVGLFGQHPVFTLMEPAHVMWTFKOSTAMMLH 91
 Db 1 SSLSMAQPEBKHVYLVSSRSGSSFFVGLFGQHPVFTLMEPAHVMWTFK----- 52

QY 92 MAVRDLIRAVFLCDMSVFPAYMEGPRRQSSLFQWENSRLCSAPACDIIPDEIIPRAH 151
 Db 53 ----- 52

QY 152 CRLCSQOPFEVVEKACRSYSHVLEKVEFNLQSLPYLPKDPNLHIVHLVDRPRAVE 211
 Db 53 -----KACRSYSHVLEKVEFNLQSLPYLPKDPNLHIVHLVDRPRAVE 98

QY 212 RSRRTKGLMIDSRVMOHQHOKLKEKEDQPTVYMOVICOSLEIKTIOSLPKALOERY 271
 Db 99 RSRRTKGLMIDSRVMOHQHOKLKEKEDQPTVYMOVICOSLEIKTIOSLPKALOERY 126

QY 272 LTVRYEDLARAPVPAQTSRMTEFVGLFPLHLQTVHNTIRGKMG 316
 Db 127 LTVRYEDLARAPVPAQTSRMTEFVGLFPLHLQTVHNTIRGKMG 171

RESULT 8
 US-10-212-933-2
 ; Sequence 2, Application US/10212933
 ; Publication No. US20030008366A1
 ; GENERAL INFORMATION:

APPLICANT: Uchimura, Kenji

APPLICANT: Muramatsu, Hideki

APPLICANT: Kadomatsu, Kenji

APPLICANT: Kannagi, Reiji

APPLICANT: Habuchi, Osami

APPLICANT: Muramatsu, Takashi

TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-

TITLE OF INVENTION: SULFOTRANSFERASE AND

FILE REFERENCE: TOYAMA1.001AUS

CURRENT APPLICATION NUMBER: US/10/212,933

CURRENT FILING DATE: 2002-08-05

PRIOR APPLICATION NUMBER: US/09/263,023

PRIOR FILING DATE: 1999-03-05

PRIOR APPLICATION NUMBER: JP 10-54007

;; PRIOR FILING DATE: 1998-03-05
 ;; PRIOR APPLICATION NUMBER: JP 10-177844
 ;; PRIOR FILING DATE: 1998-06-24
 ;; NUMBER OF SEQ ID NOS: 10
 ;; SOFTWARE: FastSeq for Windows Version 3.0
 ;; SEQ ID NO 2
 ;; LENGTH: 483
 ;; TYPE: PRT
 ;; ORGANISM: Mus musculus
 ; US-10-212-933-2

Query Match 29.4%; Score 598.5; DB 15; Length 483;
 Best Local Similarity 36.8%; Pred. No. 2, 9e-53;
 Matches 134; Conservative 64; Mismatches 137; Indels 29; Gaps 6;

QY 41 ERHNVLVSSWRSGSSFFVGLFGQHPVFTLMEPAHVMWTFKOSTAMMLHVAVRDLIRA 100
 Db 116 KRGLVYFTWTRSGSSFFBELFNQNEFFVLEPVHVNQKLYPGAVSLQGAARMLSA 175

QY 101 VFCDMSVFPAYMEGPRRQSSSLFQWENSRLCSAPACDIIPDEIIPRAHCRLL- 155
 Db 176 LYRCDLSVFLYSPASGGRNLTTLGIFGAATKRVCCSSPLCPAY-RKEVVGILVDRVCK 234

QY 156 -CSQOPFEVVEKACRSYSHVLEKVEFNLQSLPYLPKDPNLHIVHLVDRPRAVRSR 214
 Db 235 KCPQRLARPEECRKRTVVIKGVRFVDAVLAPLKDPLDKVHLVDRPRAVRSR 294

QY 215 ERTKGLMIDSRVMOHQHOKLKEKEDQ-----PYV---VMQVICOS 252
 Db 295 IRRHOLLINESLQVNASRDPRAHMPLEAGHKLKAKEGNGPADYHALGAMETICS 354

QY 253 QLEIYTIOSLPALDERILVYEDLARAPVPAQTSRMTEFVGLFPLHLQTVHNTIRG 312
 Db 355 MARTLOTALQPPMLQGHVLYRYEDLVDPVTLRYVDVDFGLVSPMEQFALNMISG 414

QY 313 KGMGDHAFHNRDALNVSOAWMSLPYEKVSRLQKACGDAMMLGRRVRSQEQBNLL 372
 Db 415 SGSSKPEFVYSARNATQANAMNTALTPOIKQVEEFCYOPNAVLDGERVNSPEEVKIDS 474

QY 373 LDLL 376
 Db 475 KTL 478

RESULT 9
 US-10-212-933-4
 ; Sequence 4, Application US/10212933
 ; Publication No. US20030008366A1
 ; GENERAL INFORMATION:

APPLICANT: Uchimura, Kenji

APPLICANT: Muramatsu, Hideki

APPLICANT: Kannagi, Reiji

APPLICANT: Habuchi, Osami

APPLICANT: Muramatsu, Takashi

TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-

TITLE OF INVENTION: SULFOTRANSFERASE AND

FILE REFERENCE: TOYAMA1.001AUS

CURRENT APPLICATION NUMBER: US/10/212,933

CURRENT FILING DATE: 2002-08-05

PRIOR APPLICATION NUMBER: US/09/263,023

PRIOR FILING DATE: 1999-03-05

PRIOR APPLICATION NUMBER: JP 10-54007

PRIOR FILING DATE: 1998-03-05

PRIOR APPLICATION NUMBER: JP 10-177844

PRIOR FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 10

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 4

LENGTH: 484

TYPE: PRT

ORGANISM: Homo sapiens

US-10-212-933-4

Query Match	28.8%;	Score 587.5;	DB 15;	Length 484;
Best Local Similarity	36.0%;	Pred. NO. 4e-52;		
Matches 131;	Conservative 67;	Mismatches 137;	Indels 29;	Gaps 6;

```

QY      41 ERNHNVLVLSMNRSGSSEFVQLFGQHPDVEYLMEPRMHNWMTFQSTAMLMHVRDLIRA 100
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      117 KRHMVLYVFTTMSGSSSEFGEELFNQNPBEVFLTEPRMHWQKLYLPEDAVSLQAGARDMLSA 176
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      101 VFLCDMSVDFAYMEPRPQRS---SLFQWENSRALCSAPACDIIPODEITPRACHRL- 155
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      177 LYRCDSYFQVLYSPAGSGRNLTLLTGIGAAITNKVYCSPLCPAY-KREYGVLDVDHYCK 235
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      156 -CSQGFVEYERKACRSYSHVYLKEYRFFNLOSLYLKDPSLNIHYLVNDPRAVFRSR 214
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      236 KCPQQLARFEECECRKRYTLTKGVDFDVAVLALTRDPLDLKVLHVLVDPRVAVASR 295
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      215 ERFGGLMIDSRIVM-----GQHEQKLKE-----DQRY---VWQYVCS 25
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      296 IRRHRLIRESLQVYNSRDPRAHRMRFLEAGCHKLGAKKEGGVGPADYHALGAEVICS 355
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      253 QLEIYKTIOSLPKALOERYLVREYEDLARAPAYQTSRMKEEYFVGEFLPHLOTWHTTGG 312
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      356 MATTLQTLQDPDWLQGHLYVRVEDYVGDPRYKTLIRRYDVLGVLLVSEMQGFALNMTSG 415
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      313 KGMGDHAFHTNADALNLSQANRWSLPYEKVSRLQKACGDAMNLGVRHNSQEQQBNLL 372
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db      416 SGSSSKRPFYVASARNQATQANMRATLTFOQIKOVEEFCYCPMAVLYGERVNSPEVYKDS 475
      : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      373 LDLL 376
      : :
Db      476 KTL 479

```

RESULT 10
US-09-833-790-255

```

Sequence 255, Application US/09833790
Patent No. US20020068288A1
GENERAL INFORMATION:
APPLICANT: Lodes, Michael J.
APPLICANT: Wang, Tongtong
APPLICANT: Secrist, Heather
APPLICANT: Mohamath, Raodoh
APPLICANT: Indrias, Carol Y.
APPLICANT: Fan, Liqun
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.512
CURRENT APPLICATION NUMBER: US/09/833,790
CURRENT FILING DATE: 2001-04-11
NUMBER OF SEQ ID NOS: 440
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 255
LENGTH: 531
TYPE: PRF
ORGANISM: Homo sapiens
US-09-833-790-255

```

Query Match	28.8%	Score 587.5	DB 9	Length 531
Best Local Similarity	36.0%	Pred. No. 4.6e-52		
Matches 131	Conservative 67	Mismatches 137	Indels 29	Gaps 6

QY	41	ERMHVLVSSRRSSSFVGOLEGGHPDVFYLMERAHNMVFKEKSTANWMLHAAYVDLIRA	100
		: : : : : : : : :	
Db	164	KRHMMVFTWRRSSSFEGLEFNONPEVFLEYEPWHWQKLYLPGDAVSLGGAARDMLSA	223
		: : : : : : : : :	
QY	101	VFLDSSFYADAYMEBGRROS----SLFQWENSRAICGAPACDIIIPDEIIIRARCRLL-	155
		: : : : : : : : : : :	
Db	224	LYRCDLSYFQLYSPAGSGGNLTTLTGIGAATNKVKVCSPLCPAY-RKEVGLVDDRYCK	282
		: : : : : : : : : :	
QY	156	-CSQPFVEVKEACRSYSHVYLKKEVRFNLSLYLLNDPSLNLNIYHLYRPRAYVRSR	214
		: : : : : : : : :	

```

Db      283 KCPORLAREEBCRCRFTLVIKGVRFVDVAIAPLBDLPALDKVTHLVDRDPAVASR 342
Qy      215 EIRKGDLMIDSRVY-----GHEBKLEK-----DQRY---VWQVICS 252
Db      343 IRRHGILIRSLQVVSRODPAHRMRFLEAAGHKLGAKEGVGGPADNALGAEVICS 402
Qy      253 QLSIYKTIOSLPKALQERYLLVREYEDLARAIPAOTSRAVEFVGLFPLHLOTWHTTG 312
Db      403 MATTLQATADQPRMLOGHLYLVRYEDVDQPKTLIRYRDYDGVLLVSEMQOFAIMTSG 462
Qy      313 KGMCDHAFHTNARDALNVSOAMRWSLPIEKSRLQACGDAMNLGRIHNSDEQRNL 372
Db      463 SGSSSRKPFVVASARNATQAAANMRTALTFOQIKOVEFCYOPMAVLGERVNSPEVKDS 522
Qy      373 LDLL 376
Db      523 KTL 526

```

RESULT 11
US-09-927

```

? Sequence 6, Application no./09927602
? Patent No. US20020061562A1
?
? GENERAL INFORMATION:
?
? APPLICANT: Fukuda, Michiko N.
?
? APPLICANT: Aama, Tomoya O.
?
? TITLE OF INVENTION: Methods of Treating Macular Cornea)
?
? TITLE OF INVENTION: Dystrophy
?
? FILE REFERENCE: P-147 4852
?
? CURRENT APPLICATION NUMBER: US/09/927,602
?
? CURRENT FILING DATE: 2001-08-09
?
? PRIOR APPLICATION NUMBER: US 09/638,211
?
? PRIOR FILING DATE: 2000-08-11
?
? NUMBER OF SEQ ID NOS: 38
?
? SOFTWARE: FastSeq for Windows Version 4.0
?
? SEQ ID NO 6
?
? LENGTH: 169
?
? TYPE: PRT
?
? ORGANISM: Homo Sapien
?
? US-09-927-602-6

```

Query Match	25.5%	Score 519.5;	DB 9;	Length 169;
Best Local Similarity	41.1%;	Pred. No. 1e-45;		
Matches 113;	Conservative 15;	Mismatches 34;	Indels 113;	Gaps 3

```

QY      42 RMHVLVLSMSRSGSSSPVGLPGOHDPVFLMEPRANHHMMFTKOSTAMHLLHNAVDDLRAV 100
      1:|||||
Db      8 RAVHVLVLSMSRSGSSSPVGLPGOHDPVFLMEPRANHHMMFT 47
QY      102 FLCDMSVFDAIYMEPGFRROSSLSLEFOWENSRALCSAPACDIIPODEIIPRACRLCSQOPF 161
      48 ----- 47
Db      162 EYVEKACRSYSHVYLKEVRFENLQSLYPLLPKPSNLHLYLVDRPRAVFRSRETKGDL 221
      1:|||||
Db      48 --LSEKCRSYSHVYLKEVRFENLQSLYPLLPKPSNLHLYLVDRPRAVFRSRETKGDL 105
QY      222 MDSRIYMGQHBEQKLKREDQPYYYVQVYICQSLQLEIKYTKIQLSPALQERYLLVYEDDLAR 281
      1:|||||
Db      106 ARDNEEA-----TLKP-RPFLRGRYRLVAFEDLAR 134
QY      282 APVAQTSRMVEFYVLEFLPHLQTYVHNITRGKMG 316
      1:|||||
Db      135 EPFLAEIRALAFPGSLTPQLEAMVHNITHGSGPG 169

```

RESULT 12

US-09-921-602-1
; Sequence 7, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:06:37 ; Search time 6872.26 Seconds

(Without alignments)
11465.210 Million cell updates/sec

Title: US-09-645-078-3

Perfect score: 1926

Sequence: 1 gggcactactactactctg.....ccctgcacatcttcttaag 1926

Scoring table: IDENTITY_NIC

Gapop 10.0 , Gapept 1.0

Searched: 2888711 seqs, 2045481386 residues

Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : GenBdb1.*

1: gb_ba.*

2: gb_htg.*

3: gb_in.*

4: gb_ov.*

5: gb_ov.*

6: gb_ov.*

7: gb_ov.*

8: gb_ov.*

9: gb_ov.*

10: gb_ov.*

11: gb_ov.*

12: gb_ov.*

13: gb_ov.*

14: gb_ov.*

15: gb_ov.*

16: gb_ov.*

17: gb_ov.*

18: gb_ov.*

19: gb_ov.*

20: gb_ov.*

21: gb_ov.*

22: gb_ov.*

23: gb_ov.*

24: gb_ov.*

25: gb_ov.*

26: gb_ov.*

27: gb_ov.*

28: gb_ov.*

29: gb_ov.*

30: gb_ov.*

31: gb_ov.*

32: gb_ov.*

33: gb_ov.*

34: gb_ov.*

35: gb_ov.*

36: gb_ov.*

37: gb_ov.*

38: gb_ov.*

39: gb_ov.*

40: gb_ov.*

41: gb_ov.*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1926	100.0	1926	6 BD134773	BD134773 GLYCOSYL
2	1926	100.0	1926	10 AF131236	AF131236 Mus muscu
3	1481.2	76.9	2201	10 AF109135	AF109135 Mus muscu
C 4	1277	66.3	232745	2 AC130174	AC130174 Rattus no
C 5	1047	54.4	262679	2 AC110314	AC110314 Rattus no
C 6	753.2	39.1	183228	2 AC10547	AC10547 Homo sapi
C 7	753.2	39.1	206943	2 AC138848	AC138848 Homo sapi
C 8	751.2	39.0	1333	9 AF149783	AF149783 Homo sapi
9	743.2	38.6	1992	9 AF280088	AF280088 Homo sapi
10	743.2	38.6	2011	9 AK026635	AK026635 Homo sapi
11	743.2	38.6	2037	9 BC035282	BC035282 Homo sapi
12	741.6	38.5	1979	6 BD127258	BD127258 Homo sapi
13	741.6	38.5	1979	9 AK074746	AK074746 Homo sapi
14	741.6	38.5	2032	6 AR203335	AR203335 Sequence
15	741.6	38.5	2032	6 BD134772	BD134772 GLYCOSYL
16	741.6	38.5	2032	9 AF131235	AF131235 Homo sapi
17	717	37.2	2065	6 AR310485	AR310485 Sequence
18	468.8	24.3	877	6 BD124832	BD124832 Primer fo
19	468.8	24.3	877	6 BD126950	BD126950 primer fo
20	370.6	19.2	1647	6 AX327330	AX327330 Sequence
21	370.6	19.2	2544	9 AF219990	AF219990 Homo sapi
22	370.6	19.2	3786	9 AF280086	AF280086 Homo sapi
23	370.6	19.2	71503	9 AF219991	AF219991 Homo sapi
C 24	370.6	19.2	157358	9 AC009163	AC009163 Homo sapi
C 25	370.6	19.2	208185	2 AC009105	AC009105 Homo sapi
26	365.2	19.0	229007	2 AC095664	AC095664 Rattus no
27	365.2	19.0	245248	2 AC117869	AC117869 Rattus no
28	355.6	18.5	1740	10 AF176841	AF176841 Mus muscu
29	355.6	18.5	1989	10 AF176840	AF176840 Mus muscu
30	355.6	18.5	159170	2 AC115914	AC115914 Mus muscu
31	355.6	18.5	209198	2 AC127315	AC127315 Mus muscu
32	354	18.4	1462	9 AF176839	AF176839 Homo sapi
33	354	18.4	2170	9 AF176838	AF176838 Homo sapi
34	354	18.4	3278	9 AF246718	AF246718 Homo sapi
35	354	18.4	194832	9 AC025287	AC025287 Homo sapi
C 36	122.6	6.4	209724	2 BX322552	BX322552 Danilo rer
C 37	116	6.0	1654	10 AB008937	AB008937 Mus muscu
38	116	6.0	7515	10 AB062107	AB062107 Mus muscu
C 39	116	6.0	223267	2 AC121503	AC121503 Mus muscu
40	114	5.9	1241	9 AF090137	AF090137 Homo sapi
41	114	5.9	1458	6 EI6306	EI6306 cDNA encod
42	114	5.9	2190	6 AR211733	AR211733 Sequence
43	114	5.9	2190	6 HS065637	HS065637 Homo sapi
44	114	5.9	2415	9 AB003791	AB003791 Homo sapi
45	114	5.9	2453	9 BC028235	BC028235 Homo sapi

ALIGNMENTS

RESULT 1

LOCUS BD134773 1926 bp DNA linear PAT 18-SEP-2002

DEFINITION GLYCOSYLulfoTransferase-3.

ACCESSION BD134773

VERSION BD134773.1 GI:23229718

KEYWORDS UP 2002507409-A/2.

SOURCE Mus sp.

ORGANISM Mus sp.

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 1926)

AUTHORS Bistup,A., Rosen,S.D., Tangemann,K. and Hemmerlich,S.

TITLE GlycosylulfoTransferase-3

JOURNAL Patent: JP 2002507409-A 2 12-MAR-2002;

Pred. No. is the number of results predicted by chance to have a

COMMENT THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, SYNTAX INC
OS Mus sp. (mouse)
JP 2002507409-A/2
PD 12-MAR-2002
PF 26-FEB-1999 JP 2000537979
PR 20-MAR-1998 US 09/045284,12-NOV-1998 US 09/190911 PI
ANNETTE BISTRUP, STEVEN D ROSEN, KIRSTEN TANGEMANN, STEFAN PI
HEMMERICH
PC C12N9/10, A01K67/027, A61K45/00, A61K48/00, A61P29/00, A61P37/06,
PC A61P43/00,
PC C12N1/15, C12N1/21, C12N5/10, C12N15/09, C12Q1/48, C12N5/00, C12N15/
PC 00
CC Glycosylsulfotransferase-3
FH key Location/Qualifiers
FT source 1.1926
FT Location/Qualifiers
1.1926 /organism="Mus sp. (mouse)".
BASE COUNT 426 a 520 c 484 g 496 t
ORIGIN
Query Match 100.0%; Score 1926; DB 6; Length 1926;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1926; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1 GGGCATCTAATCTTACACTGTGTGACAGACAGAGCTTGTGCTACAAAGGCGACACTGT 60
1 GGGCATCTAATCTTACACTGTGTGACAGACAGAGCTTGTGCTACAAAGGCGACACTGT 60
61 GTGAGGGGTGAGAAAGGTGTGGGTGTGGCAGAACTCCCTATAGTATTAATGTCTG 120
61 GTGAGGGGTGAGAAAGGTGTGGGTGTGGCAGAACTCCCTATAGTATTAATGTCTG 120
121 GGTAGGATATCTCGGTGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 180
121 GGTAGGATATCTCGGTGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 180
121 GGTAGGATATCTCGGTGTGTGATGATGATGATGATGATGATGATGATGATGATGAT 180
181 TGTACATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
181 TGTACATAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 240
241 CAATATCTGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
241 CAATATCTGACAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
301 CTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
301 CTGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 360
361 TCCCTCTGCTGCTCTTCAAGGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
361 TCCCTCTGCTGCTCTTCAAGGTCTTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 420
421 AAGGAGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
421 AAGGAGGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 480
481 TGTGCTGACAGAGACCTTTTCCAGAGAGAGAGATGATGATGATGATGATGATGATGAT 540
481 TGTGCTGACAGAGACCTTTTCCAGAGAGAGAGATGATGATGATGATGATGATGATGAT 540
541 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
541 TGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 600
601 ATGTGCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
601 ATGTGCTTACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 660
661 CCTGAGAGCTGACATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720

DB 661 CCTGAGAGCTGACATGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720
DB 721 GGTCTTTATATGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
DB 721 GGTCTTTATATGCTTACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 780
DB 781 AGCAAAAGCGGGGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 781 AGCAAAAGCGGGGCGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 840
DB 841 GCTACCCCAAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 841 GCTACCCCAAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 901 CCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
DB 901 CCTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 960
DB 961 TGTATCTACTACTGACAGAGACCTTCCCTGACCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB 961 TGTATCTACTACTGACAGAGACCTTCCCTGACCTGCTGCTGCTGCTGCTGCTGCTGCT 1020
DB 1021 CCGGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
DB 1021 CCGGGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1080
DB 1081 TTGTGCTAGGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
DB 1081 TTGTGCTAGGAGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1140
DB 1141 AGATCATCTGCAAAAGCCAGGTGACATGATGATGATGATGATGATGATGATGATGAT 1200
DB 1141 AGATCATCTGCAAAAGCCAGGTGACATGATGATGATGATGATGATGATGATGATGAT 1200
DB 1201 TGTACAGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
DB 1201 TGTACAGAGCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1260
DB 1261 CGACAGAGCTATATTAATTTGTGGGTGATGATGATGATGATGATGATGATGATGATGAT 1320
DB 1261 CGACAGAGCTATATTAATTTGTGGGTGATGATGATGATGATGATGATGATGATGATGAT 1320
DB 1321 ACAATGTCACCGCGGCAAGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1321 ACAATGTCACCGCGGCAAGGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1380
DB 1381 CCTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
DB 1381 CCTCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1440
DB 1441 AAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
DB 1441 AAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1500
DB 1501 AACAAAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
DB 1501 AACAAAGCACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1560
DB 1561 GAGAAGGTTAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
DB 1561 GAGAAGGTTAAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1620
DB 1621 ACAGAAAGCTTAAGTATTAACAAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
DB 1621 ACAGAAAGCTTAAGTATTAACAAAGTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1680
DB 1681 GGTCTTTGTCTATATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 1740
DB 1681 GGTCTTTGTCTATATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCTATGCT 1740
DB 1741 ATCTTTGTCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800
DB 1741 ATCTTTGTCTTGAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 1800

Db 1741 ATCTTCTGCTTGAAAACTATTAGAACCTTTAAGACAGCCCTTTGACTGTCAAGCAAG 1800

Qy 1801 ACCTTTCTTGAACCTTGCCCTTCTTACTGTGATCTTGAGACACGCGCTGAGGCA 1860

Db 1801 ACTTTCTTGAACCTTGCCCTTCTTACTGTGATCTTGAGACACGCGCTGAGGCA 1860

Qy 1861 TACTGACACAGCAACAGCATCTGTGAGTGTGTCTGTAAACCTCCCTGTCAACATCTT 1920

Db 1861 TACTGACACAGCAACAGCATCTGTGAGTGTGTCTGTAAACCTCCCTGTCAACATCTT 1920

Qy 1921 TCTAAG 1926

Db 1921 TCTAAG 1926

RESULT 2

AF131236 1926 bp DNA linear ROD 10-AUG-1999

LOCUS AF131236

DEFINITION Mus musculus N-acetylglucosamine 6-O-sulfotransferase gene,

ACCESSION AF131236

VERSION AF131236.1 GI:4927115

KEYWORDS

SOURCE Mus musculus (house mouse)

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

PUBMED

REFERENCE

AUTHORS

TITLE

JOURNAL

Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience,

3401 Hillview Avenue, Palo Alto, CA 94304, USA

Location/Qualifiers

1. 1926

/organism="Mus musculus"

/mol_type="genomic DNA"

/strain="C57BL/6"

/db_xref="taxon:10090"

/cell_type="high endothelial"

/tissue_type="tonsil"

<405. >1571

/product="N-acetylglucosamine 6-O-sulfotransferase"

405. 1571

/function="biosynthesis of endothelial ligands for

L-selectin"

/note="HEC-G1CMAC-6-ST"

/codon_start=1

/product="N-acetylglucosamine 6-O-sulfotransferase"

/protein_id="A033016.1"

/db_xref="GI:4927116"

/translation="MMLKRGRLMFLGSOVIVVAFIHSVHRHLSQRESRRPVH

LVSWSRSGSSFYGOLEFGHPDVEFYLMPEAMHWMTSTAMKLHAAVDLRSVL

CDMSVPAVYNNPGRKOSLFEWOSRALSAPYCDPEPHETSPHCKLSCGOPE

DYVEKARSHGFVFLKEVRLSLOALPLTDPSLNHHVHIVADPPAVRSREHTTI

ELWDSHIVAGFHLKEETKEEDOPVYAKIITCKSSQVYVKAICITLPEALDRIYLRIE

DIVARPLAQTTRIKRYGLDFLPHLQTVNVTNRKGKGQHAFTNARNLNVSAWR

WSLPEYVSQLODAGCAMDLGLIYVRSQEQGNLSLDSLSSHLIGOVFREG"

BASE COUNT 426 a 520 c 484 g 496 t

ORIGIN

Query Match 100.0%; Score 1926; DB 10; Length 1926;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1926; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 GGCATCTAATCTTACCTTGCTGAGACAGCAAGCTTTGCTTACAAGGCCACACTCT 60

Db 1 GGCATCTAATCTTACCTTGCTGAGACAGCAAGCTTTGCTTACAAGGCCACACTCT 60

Qy 61 GTCAGGGGTGTAGAAGGTGTGGGTGTGGCAGAACTCCCTATAGTATTAATGTGCTG 120

Db 61 GTCAGGGGTGTAGAAGGTGTGGGTGTGGCAGAACTCCCTATAGTATTAATGTGCTG 120

Qy 121 GGTAGATATTTCTGGTGTGTTGATGATGAGAAAGCCAGAGGGTAGTTTAAAGACT 180

Db 121 GGTAGATATTTCTGGTGTGTTGATGATGAGAAAGCCAGAGGGTAGTTTAAAGACT 180

Qy 181 TGTAACTAGAAATCAGTATCCAAATTAAGGCCAGAAATTCCTTGCAGAGGATCTGGA 240

Db 181 TGTAACTAGAAATCAGTATCCAAATTAAGGCCAGAAATTCCTTGCAGAGGATCTGGA 240

Qy 241 CAATTAATCTGAGGAATGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 300

Db 241 CAATTAATCTGAGGAATGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTTTGTGTT 300

Qy 301 CTTGATGAGGAATCCAGAAAGCCGAAAGTATGCTGTAACTAATCTAGCCCA 360

Db 301 CTTGATGAGGAATCCAGAAAGCCGAAAGTATGCTGTAACTAATCTAGCCCA 360

Qy 361 TCCCTCTGCTTGTCTTTTCAAGGCTTCTCTTCCGAGAGATGATGCTGTGAAGA 420

Db 361 TCCCTCTGCTTGTCTTTTCAAGGCTTCTCTTCCGAGAGATGATGCTGTGAAGA 420

Qy 421 AAGGAGGCTGCTATGTTCTCGGTTCCAGGTCATGTTGATGCTTTCATCCATA 480

Db 421 AAGGAGGCTGCTATGTTCTCGGTTCCAGGTCATGTTGATGCTTTCATCCATA 480

Qy 481 TGTCCGTCACAGACACTTTCACAGAGAGAGAGTCCAGAGGCCGCTGATGCTGCG 540

Db 481 TGTCCGTCACAGACACTTTCACAGAGAGAGAGTCCAGAGGCCGCTGATGCTGCG 540

Qy 541 TGTCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600

Db 541 TGTCTGTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 600

Qy 601 ATGTGTTCTACCTGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660

Db 601 ATGTGTTCTACCTGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 660

Qy 661 CCTGGAAGCTGACATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720

Db 661 CCTGGAAGCTGACATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 720

Qy 721 GCGTCTTGAATGCTTACATGAAACCCAGGCCCCGAAACAGTCCAGCTTCCAGTGGG 780

Db 721 GCGTCTTGAATGCTTACATGAAACCCAGGCCCCGAAACAGTCCAGCTTCCAGTGGG 780

Qy 781 AGCAAGAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840

Db 781 AGCAAGAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840

Qy 841 GCTACCAAGCACTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900

Db 841 GCTACCAAGCACTGCAAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 900

Qy 901 CCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960

Db 901 CCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960

Qy 961 TCTATCACTACTACAGGACCTTCCCTCAACCTGACAGCTGCTGCTGCTGCTGCTGCTG 1020

Db 961 TCTATCACTACTACAGGACCTTCCCTCAACCTGACAGCTGCTGCTGCTGCTGCTGCTG 1020

Qy 1021 CCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080

Db 1021 CCGGCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1080

OY 1081 TTGTCGTAAGGAGCATTTGGAAGATGCAAGAGAGAACCCAGCCCTTTATGCCATGA 1140
 DB 1081 TTGTCGTAAGGAGCATTTGGAAGATGCAAGAGAGAACCCAGCCCTTTATGCCATGA 1140
 OY 1141 AGATCATCTGCAAAAGCCAGGTGACATAGTCAAGGCCATCCAAACCCCTCTGAAGCTC 1200
 DB 1141 AGATCATCTGCAAAAGCCAGGTGACATAGTCAAGGCCATCCAAACCCCTCTGAAGCTC 1200
 OY 1201 TGCAGACGGCTACCTGTTCTGAGATGAGAGACCTGGTCCGAGCCCTGGCCAGCA 1260
 DB 1201 TGCAGACGGCTACCTGTTCTGAGATGAGAGACCTGGTCCGAGCCCTGGCCAGCA 1260
 OY 1261 CGACAGACTATATTAATTTGGGGTGGATTTTTTGGCCCAACCTCCAAAGATGGCTT 1320
 DB 1261 CGACAGACTATATTAATTTGGGGTGGATTTTTTGGCCCAACCTCCAAAGATGGCTT 1320
 OY 1321 ACATATGTCACCCGCGGCAAGGAGCATGAGTCAAGATGCTTCCATACATAAGCAGAGACG 1380
 DB 1321 ACATATGTCACCCGCGGCAAGGAGCATGAGTCAAGATGCTTCCATACATAAGCAGAGACG 1380
 OY 1381 CCCCTAACGCTCTCTGAGCGGTGGCTGCTTACCTTACGAAAAAGTTTCCAGCTTC 1440
 DB 1381 CCCCTAACGCTCTCTGAGCGGTGGCTGCTTACCTTACGAAAAAGTTTCCAGCTTC 1440
 OY 1441 AAGATGCTGCTGAGGCTATGAGATTTCTGAGGATGCTCCAGATGCTCAAGCAAG 1500
 DB 1441 AAGATGCTGCTGAGGCTATGAGATTTCTGAGGATGCTCCAGATGCTCAAGCAAG 1500
 OY 1501 AACAGAGCAAGCTGCTGAGATGCTGCTGCTCCATGCTCCAGATGCTGAGGAGGCTTCC 1560
 DB 1501 AACAGAGCAAGCTGCTGAGATGCTGCTGCTCCATGCTCCAGATGCTGAGGAGGCTTCC 1560
 OY 1561 GAGAGGTTAAGAGGCTGCTGCTGCAACCCCTGGTTCAGCTTACGATCAATTAACGC 1620
 DB 1561 GAGAGGTTAAGAGGCTGCTGCTGCAACCCCTGGTTCAGCTTACGATCAATTAACGC 1620
 OY 1621 ACAGAGGCTTAAAGTATACCAAACTGAGTCCCTTCTCTGAGCCCAAGCAGAG 1680
 DB 1621 ACAGAGGCTTAAAGTATACCAAACTGAGTCCCTTCTCTGAGCCCAAGCAGAG 1680
 OY 1681 GGTCTTTGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 1740
 DB 1681 GGTCTTTGTCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCTATCT 1740
 OY 1741 ATCTTTCTGCTGTAATAATTAATTAAGAACCTTAAGAGCCCTTTGACCTGTCAAGCAAG 1800
 DB 1741 ATCTTTCTGCTGTAATAATTAATTAAGAACCTTAAGAGCCCTTTGACCTGTCAAGCAAG 1800
 OY 1801 ACTTCTGTAACCTTGCGCTTCTTACCTGATGATACCTTGAAGACTGGTCTGAGAGCA 1860
 DB 1801 ACTTCTGTAACCTTGCGCTTCTTACCTGATGATACCTTGAAGACTGGTCTGAGAGCA 1860
 OY 1861 TACTGAGACAGCAAAAGCATCTGTGAGATGTGTCTGTAACCTGCTGTCACATCTTT 1920
 DB 1861 TACTGAGACAGCAAAAGCATCTGTGAGATGTGTCTGTAACCTGCTGTCACATCTTT 1920
 OY 1921 TCTAAG 1926
 DB 1921 TCTAAG 1926
 RESULT 3
 AFI09155 2201 bp mRNA linear ROD 27-JUL-1999
 LOCUS Mus musculus L-selectin ligand sulfotransferase mRNA, complete cds.
 DEFINITION AFI09155
 ACCESSION AFI09155.1 GI:5596405
 VERSION AFI09155.1 GI:5596405
 KEYWORDS
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 REFERENCE 1 (bases 1 to 2201)

AUTHORS Hirooka, N., Petryniak, B., Nakayama, J., Tsuboi, S., Suzuki, M.,
 Yeh, J.-C., Izawa, D., Tanaka, T., Miyasaka, M., Lowe, J. B. and Fukuda, M.
 TITLE A novel, high endothelial venule-specific sulfotransferase
 expresses 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed
 by CD34
 JOURNAL Immunology 11 (1), 79-89 (1999)
 MEDLINE 99361934
 PUBMED 10435581
 REFERENCE 2 (bases 1 to 2201)
 AUTHORS Hirooka, N. and Fukuda, M.
 TITLE Direct Submission
 JOURNAL Submitted (23-NOV-1998) Glycobiology, The Burnham Institute, 10901
 North Torrey Pines Road, La Jolla, CA 92037, USA
 FEATURES
 source Location/Qualifiers
 1..2201
 /organism="Mus musculus"
 /mol_type="mRNA"
 /db_xref="taxon:10090"
 319..1485
 /function="directs expression of 6-sulfo sialyl Lewis x,
 an L-selectin ligand displayed by CD34"
 /note="LSSR"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase"
 /protein_id="ABD45579.1"
 /db_xref="GI:5596406"
 /translation="MMLKKGRILNLFSGQIVVALFIMSVHRHLSORESRPVH
 LVLSWRSSSFFVQDFGQHPDVFLMEAHMVMFTFSTAMKLMARLDLRSVFL
 CDSMSVEDAMNRPKROSLPQWEOBALSCPDVDFPHEISPRKFKLGGQPF
 DMVEKRSRSHGVVILKEVRSLSQALYPLPSLNTHVYHLYVRPFRREHRTT
 ELVDSHYLGGHLETKEDDPTYAMKLTICQSYDIKAIOTLEBALQRLFLATE
 DLVRAPLAQTFLRYFVGLDFPHLQTVHNTVRKMGQAHFNARNAVLVSAMR
 WSLPEKXSQLDADCEAMDLGLYQVRSQDQBNLSIDLSSSHILGQVREG"
 BASE COUNT 537 a 629 c 534 g 501 t
 ORIGIN
 Query Match 76.9%; Score 1481.2; DB 10; Length 2201;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1503; Conservative 0; Mismatches 18; Indels 1; Gaps 1;
 OY 402 AGGATGATCTGCTGTAAGAAAGAGGAGCTGATGTTCTGGGTTCCAGGTCATCGTT 461
 DB 316 AGGATGATCTGCTGTAAGAAAGAGGAGCTGATGTTCTGGGTTCCAGGTCATCGTT 375
 OY 462 GTAGCTCTTATCATCATATGTCGCTCAGACACACCTTCCAGAGGAGAGTCCAGG 521
 DB 376 GTAGCTCTTATCATCATATGTCGCTCAGACACACCTTCCAGAGGAGAGTCCAGG 435
 OY 522 AGCCCGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 581
 DB 436 AGCCCGGTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 495
 OY 582 CTTTTCGGGACACACCGGATGTTCTGCTGATGAGACCTGCGGATGCTGATG 641
 DB 496 CTTTTCGGGACACACCGGATGTTCTGCTGATGAGACCTGCGGATGCTGATG 555
 OY 642 ACTTTCACGACGACACGCTGGAAGTGCATGCTGCTGCGGAGATCTTGCCTCC 701
 DB 556 ACTTTCACGACGACACGCTGGAAGTGCATGCTGCTGCGGAGATCTTGCCTCC 615
 OY 702 GTCCTCCTGCTGATGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 761
 DB 616 GTCCTCCTGCTGATGATGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 675
 OY 762 TCCAGACCTTTCACATGAGGAGCAAGCCGCGGCTGTGCTGACGCTGTGTGACTTC 821
 DB 676 TCCAGACCTTTCACATGAGGAGCAAGCCGCGGCTGTGCTGACGCTGTGTGACTTC 735
 OY 822 TTCCCTGCCCAGAGATCAGTCAACCAAGCATGCAAGCTGCTTGGCTGACGACGCC 881
 DB 736 TTCCCTGCCCAGAGATCAGTCAACCAAGCATGCAAGCTGCTTGGCTGACGACGCC 795
 OY 882 TTGATATGCTGAGAGAGCCCTGCGGCTCTCAAGGCTTCTGCTACTCAAGAGAGTGGCT 941

Db 796 TTGATATGATGAGAGGCGCTGCGCTCAGCGCTTCGAGTACCAAGAGAGTGCGT 855
 Qy 942 TTTCAGCTGAGAGGCGCTCTATCCACTACTCAGAGACCTTCCCTCAACCTCAGTGC 1001
 Db 856 TTTCAGCTGAGAGGCGCTCTATCCACTACTCAGAGACCTTCCCTCAACCTCAGTGC 915
 Qy 1002 GTGACACTGATCCGAGAGACCCCGGCGGTGTCCGATCCCGGAGCAGACACCATAGAA 1061
 Db 916 GTGACACTGATCCGAGAGACCCCGGCGGTGTCCGATCCCGGAGCAGACACCATAGAA 975
 Qy 1062 CTCATGTTGACACTCATATTGTGCTAGGAGCAGATTGGAAAGCATCAAGAGAAAGAC 1121
 Db 976 CTCATGTTGACACTCATATTGTGCTAGGAGCAGATTGGAAAGCATCAAGAGAAAGAC 1035
 Qy 1122 CAGCCCTATATGCGAGAGATCATGCAAAAGCCAGGTGAGCATAGTCAAGCCATC 1181
 Db 1036 CAGCCCTATATGCGAGAGATCATGCAAAAGCCAGGTGAGCATAGTCAAGCCATC 1095
 Qy 1182 CAAACCTCCCTGAGAGCTGAGAGAGAGCTGCTGCTGAGATGAGAGAGCTGTT 1241
 Db 1096 CAAACCTCCCTGAGAGCTGAGAGAGAGCTGCTGCTGAGATGAGAGAGCTGTT 1155
 Qy 1242 CGGGACACCTGGCGCCAGAGAGACCACTATATAATTGTTGGGGTTGGATTTTGGCC 1301
 Db 1156 CGGGACACCTGGCGCCAGAGAGACCACTATATAATTGTTGGGGTTGGATTTTGGCC 1215
 Qy 1302 CACCTCAAAACATGGGTTTCAATGATGACACCGCGGCAAGGAGCATGGGTGAGCAGACCTTC 1361
 Db 1216 CACCTCAAAACATGGGTTTCAATGATGACACCGCGGCAAGGAGCATGGGTGAGCAGACCTTC 1275
 Qy 1362 CATACTAAGCGCAGAGAGAGCGCTTCAAGCTCTCTGAGCGTGGCGTGGTCTTACCTTAC 1421
 Db 1276 CATACTAAGCGCAGAGAGAGCGCTTCAAGCTCTCTGAGCGTGGCGTGGTCTTACCTTAC 1335
 Qy 1422 GAAAGGTTTCCAGCTTCAAGATGCTGCGGTAGAGCTATGATTTGCTGGATPACCTC 1481
 Db 1336 GAAAGGTTTCCAGCTTCAAGATGCTGCGGTAGAGCTATGATTTGCTGGATPACCTC 1395
 Qy 1482 CAGGTAGATCTCAACAAGAACAGAGACCTGTCCTGATCTTCTGCTCTCCAT 1541
 Db 1396 CAGGTAGATCTCAACAAGAACAGAGACCTGTCCTGATCTTCTGCTCTCCAT 1455
 Qy 1542 ATCTTGGGGAGGTCTTCCGAGAGGTTAAGAGAGTGTCTGCACCCCTTGTCCAGC 1601
 Db 1456 ATCTTGGGGAGGTCTTCCGAGAGGTTAAGAGAGTGTCTGCACCCCTTGTCCAGC 1515
 Qy 1602 CTTAGTACACTTAAAGCAGAGACCTTAGGTATTAACCAACTGATGCCCTTCT 1661
 Db 1516 CTTAGTACACTTAAAGCAGAGACCTTAGGTATTAACCAACTGATGCCCTTCT 1575
 Qy 1662 CCTGAGCCCAAGAGAGAGGCTTTGTTGTTATCATGTTACCTTACCAAGTGAAGC 1721
 Db 1576 CCTGAGCCCAAGAGAGAGGCTTTGTTGTTATCATGTTACCTTACCAAGTGAAGC 1635
 Qy 1722 TAAAGAGCCAAAGAAAGATCTTCTGCTTGAAGTACTTAGGAACCTTAAAGCAGCC 1781
 Db 1636 TAAAGAGCCAAAGAAAGATCTTCTGCTTGAAGTACTTAGGAACCTTAAAGCAGCC 1695
 Qy 1782 CTTTACCTGTCAGAGAGACTTCTTGTAACTCTTGGCCTTCTTACCTGTCATACCTTG 1841
 Db 1696 CTTTACCTGTCAGAGAGACTTCTTGTAACTCTTGGCCTTCTTACCTGTCATACCTTG 1755
 Qy 1842 GAGAGCT-CGGTGGAGAGAGATCATGAGACAGCAAGAAAGATCTGTCGAGTGTCTGTA 1900
 Db 1756 GAGAGCTAGGCTGTGAGAGGCTCTAGTGGACAGCAAGAAAGATCATGAGTGTCTGTA 1815
 Qy 1901 AACCTCCCTGTACACATCTTTC 1922
 Db 1816 AACCTCCCTGTACACATCTTTC 1837

RESULT 4

AC130174/c
 LOCUS 232745 bp DNA linear HTG 13-NOV-2002
 DEFINITION Rattus norvegicus clone CH230-14011, WORKING DRAFT SEQUENCE, 3
 AC130174
 accession AC130174.5 GI:24635752
 version HTG; HTGS; PHASE1; HTGS; DRAFT; HTGS; FULLTOP.
 keywords Rattus norvegicus (Norway rat)
 ORGNAMISM Rattus norvegicus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
 Rattus.
 1 (bases 1 to 232745)
 Mueny,D,Marle, Metzker,M,lee, Abramson,S, Adams,C, Alder,J,
 Allen,C, Allen,H, Alshrocks,S, Anlin,A, Anquiano,D,
 Anyalbechl,V, Aoyagi,A, Ayodeji,M, Baca,E, Baden,H,
 Baldwin,D, Bandaranaike,D, Barber,M, Bainsstead,M, Benahmed,F,
 Bismalo,K, Blair,J, Blankenburg,K, Blyth,P, Brown,M,
 Bryant,N, Buhay,C, Burch,P, Burrell,K, Calderon,E,
 Cardenas,V, Carter,K, Cavazos,I, Ceasar,H, Center,A,
 Chacko,J, Chavez,D, Chen,G, Chen,R, Chen,Y, Chen,Z, Chu,J,
 Cleveland,C, Cockrell,R, Cox,C, Coyle,M, Cree,A, D'Souza,L,
 Davila,M,L, Davis,C, Davy-Carroll,L, De Anda,C, Dederich,D,
 Delgado,O, Denson,S, Deramo,C, Ding,Y, Dinh,H, Diya,K,
 Diaper,H, Dugan-Rocha,S, Dunn,A, Dugdin,K, Duval,B, Eaves,K,
 Egan,A, Escotto,M, Eugene,C, Evans,C,A, Falls,T, Fan,G,
 Fernandez,S, Finley,M, Flagg,N, Forbes,L, Foster,M, Foster,P,
 Fraser,C,M, Gabisi,A, Gante,R, Garcia,A, Garner,T, Garza,M,
 Gedregeorgis,E, Geer,K, Gill,R, Grady,M, Guerra,M, Guevara,W,
 Gunaratne,P, Haaland,W, Hamill,C, Hamilton,K, Hamilton,K,
 Harvey,Y, Havlak,P, Hawes,A, Henderson,N, Hernandez,J,
 Hernandez,R, Hines,S, Hladun,S,L, Hodgson,A, Hognes,M,
 Hollins,B, Howells,S, Huliy,S, Hume,J, Idlebird,D, Jackson,A,
 Jackson,L, Jacob,L, Jiang,H, Johnson,B, Johnson,R, Jollivet,A,
 Karpachy,S, Kelly,S, Kelly,S, Khan,Z, King,K, Kova,C,
 Kowis,C, Kraft,C,L, Lebow,H, Levan,J, Lewis,L, Li,Z, Liu,J,
 Liu,J, Liu,W, Liu,Y, London,P, Longacre,S, Lopez,J,
 Lorensuwa,L, Louisedge,H, Lozardo,R,J, Lu,X, Ma,J,
 Maheshwari,M, Mahindratne,M, Mahmoud,M, Malloy,K, Mangum,A,
 Mangum,B, Mapua,P, Martin,K, Martin,R, Matinez,E,
 Mawhney,S, McLeod,M,P, McNeill,T,Z, Meenen,E,
 Mitosavljevic,A, Miner,G, Minja,E, Montemayor,J, Moore,S,
 Morgan,M, Morris,K, Morris,S, Muidasa,M, Murphy,M, Nair,L,
 Nankervis,C, Neal,D, Newton,N, Nguyen,N, Norris,S,
 Nwokenleh,O, Okwunu,G, Olarpunsaogon,A, Pal,S, Parks,K,
 Pasternak,S, Paul,H, Perez,A, Perez,L, Pfannkuch,C,
 Piopper,F, Polidexter,A, Popovic,D, Primus,E, Pu,L,L,
 Puzo,M, Quiroz,J, Rachlin,E, Reeves,K, Regier,M,A, Reigh,R,
 Reilly,B, Reilly,M, Ren,Y, Reuter,M, Richards,S, Riggs,F,
 Rives,C, Rodkey,T, Rojas,A, Rose,M, Rose,R, Ruiz,S,J,
 Sanders,W, Saverly,G, Scherer,S, Scott,G, Shatman,S, Shen,H,
 Shetty,J, Shvartsbeyn,A, Sisson,I, Sitter,C,D, Smeja,D,
 Sneed,A, Sodergren,E, Song,X,Z, Soreller,R, Sosa,J,
 Steimle,M, Strong,R, Sutton,A, Swalek,A, Taber,P, Taylor,C,
 Taylor,T, Thomas,N, Thomas,S, Tingey,A, Trejos,Z, Usmani,K,
 Valas,R, Vera,V, Villasa,D, Waldron,L, Walker,B, Wang,J,
 Wang,O, Wang,S, Warren,J, Warren,R, Wei,X, White,K,
 Williams,G, Willson,R, Wlecyk,R, Wooden,H, Morley,K,
 Wright,D, Wright,R, Wu,D, Yakub,S, Yen,J, Yoon,L, Yoon,V,
 Yu,F, Zhang,J, Zhou,J, Zhou,X, Zhao,S, Dunn,D, von
 Niederhausern,A, Weiss,R, Smith,D,R, Holt,R,A, Smith,H,O,
 Weinstock,G, and Gibbs,R.A.
 Direct Submission
 Unpublished
 2 (bases 1 to 232745)
 Morley,K,C.
 Direct Submission
 Submitted (08-AUG-2002) Human Genome Sequencing Center, Department
 of Molecular and Human Genetics, Baylor College of Medicine, One
 Baylor Plaza, Houston, TX 77030, USA
 3 (bases 1 to 232745)
 Rat Genome Sequencing Consortium.
 Direct Submission

JOURNAL

COMMENT

Submitted (13-NOV-2002) Human Genome Sequencing Center, Department of Molecular and Human Genetics, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030, USA
 On Nov 6, 2002 this sequence version replaced 91:23322193.
 The sequence in this assembly is a combination of BAC based reads and whole genome shotgun sequencing reads assembled using Atlas (<http://www.hgsc.bcm.tmc.edu/projects/atlas/>). Each contig described in the feature table below represents a scaffold in the Atlas assembly (a 'contig-scaffold'). Within each contig-scaffold, individual sequence contigs are ordered and oriented, and separated by sized gaps filled with Ns to the estimated size. The sequence may extend beyond the ends of the clone and there may be sequence contigs within a contig-scaffold that consist entirely of whole genome shotgun sequence reads. Both end sequences and whole genome shotgun sequence only contigs will be indicated in the feature table.

----- Genome Center

Center: Baylor College of Medicine

Center code: BCM

Web site: <http://www.hgsc.bcm.tmc.edu/>

Contact: hgsc-help@bcm.tmc.edu

----- Project Information

Center project name: GEKI

Center clone name: CH230-14J11

----- Summary Statistics

Assembly program: Phrap; version 0.990329

Consensus quality: 221656 bases at least Q40

Consensus quality: 224477 bases at least Q30

Consensus quality: 226385 bases at least Q20

Estimated insert size: 228311; sum-of-contigs estimation

Quality coverage: 7x in Q20 bases; sum-of-contigs estimation

 * NOTE: Estimated insert size may differ from sequence length
 * (see http://www.hgsc.bcm.tmc.edu/docs/genbank/draft_data.html).
 * NOTE: This is a 'working draft' sequence. It currently
 * consists of 3 contigs. The true order of the pieces
 * is not known and their order in this sequence record is
 * arbitrary. Gaps between the contigs are represented as
 * runs of N, but the exact sizes of the gaps are unknown.
 * This record will be updated with the finished sequence
 * as soon as it is available and the accession number will
 * be preserved.

* 1 220755: contig of 220755 bp in length
 * 220756 220855: gap of unknown length
 * 220856 231390: contig of 10535 bp in length
 * 231391 231490: gap of unknown length
 * 231491 232745: contig of 1255 bp in length.

----- Location/Qualifiers

1. 232745

/organism="Rattus norvegicus"

/mol_type="genomic DNA"

/db_xref="taxon:10116"

/clone="CH230-14J11"

1. 1118

/note="wgs.contig"

misc_feature 220856..222264

misc_feature 220856..222264

BASE COUNT 60143 a 50263 c 51032 g 65918 t 5389 others

ORIGIN

Query Match 66.3%; Score 1277; DB 2; Length 232745;

Best Local Similarity 87.2%; Pred. No. 0;

Matches 1496; Conservative 0; Mismatches 175; Indels 44; Gaps 7;

208 AAGACCAAGATTACTTTCAGAGAGATCGACAATAATCTGACGAGATGTTTGGTT 267

DB 176753 AAGAACAATTAATCTCTCAGAGAGATCGACAATAATCTGACGAGATGTTTGGTT 176694

QY 268 TTGTTGTTGTTGTTGTTTACATTCCTGATGAGGATTCAGAGAGAGCCCA 327

DB 176693 -----TTTACATTCATGATCTTGAGAGGAGGCTTACAGAGAGCCCA 176654

QY 328 AGTAGATGCTGTACACACTACACGCCCATCCCTCTGCTCTTCAAGGTCT 387

QY 328 AGTAGATGCTGTACACACTACACGCCCATCCCTCTGCTCTTCAAGGTCT 387

DB 176653 GGGTAGATACCTATACAGCCTATACAGCCCGCTCCCTCTCTTCCCTTAAGGT-- 176596

QY 388 TCTCCCTTCGCGAGATGATGCTGTGAAGAAAGGAGGCGGATGCTCGGGT 447

DB 176595 ----CTACTTCGCGAGATGATGCTGTGAAGAAAGGAGGCGGATGCTCGGGT 176540

QY 448 CCCAGGTATCGTTGATGCTGTTCATCATATGCTCCACAGACACCTTCCACA 507

DB 176539 CCCAGGTATGATGATGCTGTTCATCATATGCTCCACAGACACCTTCCACA 176480

QY 508 GGGAGAGTCCAGAGAGGCGGCTGATGCTGTGCTGTCTTCTTCCGCGGTAGAGTCT 567

DB 176479 GGGAGAGTCCAGAGAGGCGGCTGATGCTGTGCTGTCTTCTTCCGCGGTAGAGTCT 176420

QY 568 CTTTGTGGAGACGCTTTCGGGACGACCGGATGCTTACCGATGAGAGCGCT 627

DB 176419 CTTTGTGGAGACGCTTTCGGGACGACCGGATGCTTACCGATGAGAGCGCT 176360

QY 628 GGCATGTGTGATGATCTTACACAGAGCAGCCTGAGAGCTGACATGCTGTGGG 687

DB 176359 GGCATGTGTGATGATCTTACACAGAGCAGCCTGAGAGCTGACATGCTGTGGG 176300

QY 688 ATCTTCTGCTTCCCTTCTTCTGATGATGATGATGATGATGATGATGATGATGAT 747

DB 176299 ATCTTCTGCTTCCCTTCTTCTGATGATGATGATGATGATGATGATGATGATGAT 176240

QY 748 GCGCCGGAAGAGTCCAGCCTTCTTCAAGTGGAGAGAGGCGGCTGCTCAAGCC 807

DB 176239 GCGCCGGAAGAGTCCAGCCTTCTTCAAGTGGAGAGAGGCGGCTGCTCAAGCC 176180

QY 808 CTGTGTGATCTTCTTCCCTGCGCAGAGATGATGATGATGATGATGATGATGATGAT 867

DB 176179 CTGTGTGATCTTCTTCCCTGCGCAGAGATGATGATGATGATGATGATGATGATGAT 176120

QY 868 GCGGTACAGAGCCTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 927

DB 176119 GCGGTACAGAGCCTTGTATGATGATGATGATGATGATGATGATGATGATGATGAT 176060

QY 928 TCAAGAGAGTGGTTCCTTTCACAGCCTGAGCCTGATATCACTACAGAGAGCTTCC 987

DB 176059 TCAAGAGAGTGGTTCCTTTCACAGCCTGAGCCTGATATCACTACAGAGAGCTTCC 176000

QY 988 TCAACCTGACGCTGCTGACCTGCTGACAGCCCGGCGCTGCTGCTGCTGCTGCTGCT 1047

DB 175999 TCAACCTGACGCTGCTGACCTGCTGACAGCCCGGCGCTGCTGCTGCTGCTGCTGCT 175940

QY 1048 ACACACCATTAACATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1107

DB 175939 ACACACCATTAACATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 175880

QY 1108 TCAAG 1167

DB 175879 TCAAG 175820

QY 1168 TAGTCAGAGCCTTCCAAACCTGCTGAAAGCTGCTGAGAGAGAGAGAGAGAGAGAGAG 1227

DB 175819 TAGTCAGAGCCTTCCAAACCTGCTGAAAGCTGCTGAGAGAGAGAGAGAGAGAGAGAG 175760

QY 1228 ATGAG 1287

DB 175759 ATGAG 175700

QY 1288 TGGATTTTGGCCCACTCCAAACATGAGTTTACATGTCACCCGCGCAAGGAGATG 1347

DB 175699 TGGATTTTGGCCCACTCCAAACATGAGTTTACATGTCACCCGCGCAAGGAGATG 175640

QY 1348 GTACAGATGCTTCCATTAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1407

DB 175639 GTACAGATGCTTCCATTAACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 175580

QY 1408 GGTCTTACCTTACAG 1467

DB 175580 GGTCTTACCTTACAG 175530

QY	493	GAAACCTTTCCAGAGGAGAGAGATCCAGAGAGCCCTGCATGTGCTGTGCTCTCTCT	552
Db	64577	GAACTTGTTCAGAGGAGAGAGATCCAGAGAGCCCTGTGCAGTGTGCTCTCTCTCT	64518
QY	553	GCGCGTCAGAGATCTCTTTTGTGTGGACAGCTTTTCGGGACAGCCCGAGTGTCTTACC	612
Db	64517	GCGGCTCGGGGTCTCTTTTGTGTGGACAGCTTTTGTGGACAGCCCGAGTGTCTTACC	64458
QY	613	TGATGAGCCTGCTCTGGCATGTGTGTGATGATCTTACACAGACAGCAGCCTGGAAAGCTGC	672
Db	64457	TGATGAGCCTGCTCTGGCATGTGTGTGATGATCTTACACAGACAGCAGCCTGGAAAGCTGC	64398
QY	673	ACATGGCTGTGGCGGAGATCTTCTCGTTCGCTTCTGTGTGATGATGACAGCTTTTGATG	732
Db	64397	ACATGGCTGTGGCGGAGATCTTCTCGTTCGCTTCTGTGTGATGATGACAGCTTTTGATG	64338
QY	733	CCTACATGAACCCAGAGCCCGCCGGAACAGTCCAGCCTCTTCCAGTGTGGAGCAAGAGCCGG	792
Db	64337	CCTACATGAACCCGGGCCCCCGGAAACATTCAGCCTCTTCCAGTGTGGAGCAAGAGCCGG	64278
QY	793	CCCTGTGCTCAGAGCCTGTGTGTGATCTTCCCTGTGCCACGAGATCACTACCCAGAC	852
Db	64277	CCCTGTGCTCTGTGCTATTTGTGGAGCTTCTCCCTGTACCAAGATCACTGTGCCCAAGC	64218
QY	853	ACTGACAGTGTGCTCTCGGTGTGCAGACCCCTTGTGATGTGTGGAGAGGCTCGGTCTGC	912
Db	64217	ATTGCAAGCGCGCTCTCAGTCAAGCTGCCCTTTGTGATGTGTGGAGAGGCTCGGTCTGC	64158
QY	913	ACGGCTGTGTGTACTCAAGAGAGGTGCGTTTCTCAGCCTGTGAGGCGCCTATTCACATAC	972
Db	64157	ACAGCCAGCTGTGTACTCAAGAGAGGTGCGTTTCTCAGCCTGTGAGGCGCCTGTATTCACATGC	64098
QY	973	TCAAGGACCTTTCCTCAACCTGTGCAGCTGTGTGTGACCTGTGCCGAGACCCCGGCGCTGT	1032
Db	64097	TCAAGACACCTTTCCTCAACCTGTGCAGCTGTGTGTGACCTGTGCCGAGACCCCGGCGCTGT	64038
QY	1033	TCCGATCCCGGAGGACACACACATCTGAACTCATGTGTGACAGTATATTGTGTCTAAGGCG	1092
Db	64037	TCCGATCCCGGAGGACACACACATCTGAACTCATGTGTGACAGTATATTGTGTCTAAGGCG	63978

OY	1093	AGCATTTTGGAACATCAAGAGAGAAACACAGCCATTATATGGCATTGAAGATCATCTGCA	1152		
Db	63977	AGAAATTTGGAAAATATATCAAGAGAGAAAGACCAAGCCCTTATATGCGCATGAAAGATCATCTGCA	63911		
OY	1153	AAAGCCAGGTGGACATAGTCAAGAGCCATCCAAACCCCTCCCTGAAAGCTCTGACAGACGCT	1212		
Db	63917	AAAGCCAGGTGGACATAGTCAAGAGCCATCCAAACCCCTCCCTGAAAGCTCTGACAGACGCT	63855		
OY	1213	ACCGTTCCCTGAGGTATGAGACCTGGTTGGGACACCCCTGGGCCACAGAGACCAAGCTAT	1272		
Db	63857	ACCGTTCCCTGAGGTATGAGAGACCTGGTTGGGACACCCCTGGGCCACAGAGACCAAGCTAT	63798		
OY	1273	ATAAATTTGGGGGTTGGATTTTGTGGCCACCCACCTCCAAACATGGATTACATGTACACC	1332		
Db	63797	ATAAATTTGGGGGTTGGATTTTGTGGCCACCCCTCCAAACATGGATTACATGTACACC	63738		
OY	1333	GCGGCAAGGCAATGGGTGACAGCATGCTTCATACTAAAGCCAGAAAGCCCTTCACGTC	1392		
Db	63737	GTCGCAAGGGCAATGGGTGACAGCATGCTTCATACTAAAGCCAGAAAGCCCTTCACGTC	63678		
OY	1393	CTCAGGGGTGGCGGTGGTGGCTTACCTTACGAAAGGTTCCAGTCCAAAGATGCTGGG	1452		
Db	63677	CTCAGGGCATGGCGGTGGTGGCTTACCTTACCTTATACAAAGGTTCTCAGCTTCAAGATGACTGTG	63618		
OY	1453	GTCAGGCTATGATTTGCTGGGATACCTCCAGGTGATGATCTCAACAAGAACAGGCAAC	1512		
Db	63617	GTCAGGCTATGATTTCTGGGATACCTCCAGGTGATGATCTCAACAAGAACAGGCAAC	63558		
OY	1513	TGTCCTCCGGAATCTTCTGCTCCTCCCATATCTTGGGGCAGGTCTTCCGAAAGTTAA	1572		
Db	63557	TGTCCTCCGGAATCTTCTGCTCCTCCCATATCTTGGGGCAGGTCTTCCGAAAGTTAA	63498		
OY	1573	GAGTCTGTGTGCAACCCCTGGTTCAGGCTTGTAGTACCATTTAAAGCAGACAGAGGCT--	1630		
Db	63497	GAGTCTGTGTGCAACCCCTGGTTCAGGCTTGTAGTACCATTTAAAGCAGACAGAGGCT--	63438		
OY	1631	--TAAAGTATTAACCAACTGATGAGCCCTTCTGCTCAGCCCCAAGAGGAGGCTTTTG	1688		
Db	63437	AGTAAAGTATTAACCAACTGATGAGCCCATTTATCTCAGTCCCAAGACAGAGGAGTCTCT--	63379		
OY	1689	TGTCATATACATGATGCTACCCCTACACTGAGCCTAAAGAGCCAAAGAAAGATATCTTCT	1748		
Db	63378	--TTTATATACACAGGTCTA--TACGGCCAAAGCCTAAAGAGCCAAAGAAAGATATCTTCT	63324		
OY	1749	GTCCTGAAATATAC	1761		
Db	63323	GTCCTGAAATATAC	63311		
RESULT 6	AC010547	183228 bp	DNA	linear	PRI 29-SEP-2001
LOCUS	AC010547				
DEFINITION	Homo sapiens chromosome 16 clone R11-510M2, complete sequence.				
ACCESSION	AC010547				
VERSION	AC010547.9				
KEYWORDS	HTG.				
SOURCE	Homo sapiens (human)				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.				
AUTHORS	1 (bases 1 to 183228)				
TITLE	DOE Joint Genome Institute.				
JOURNAL	Sequencing of Human Chromosome 16				
REFERENCE	Unpublished				
AUTHORS	2 (bases 1 to 183228)				
TITLE	DOE Joint Genome Institute.				
JOURNAL	Direct Submission				
REFERENCE	Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA				
AUTHORS	3 (bases 1 to 183228)				
TITLE	DOE Joint Genome Institute.				
JOURNAL	Direct Submission				

DEFINITION Homo sapiens L-selectin ligand sulfotransferase mRNA, complete cds.
 ACCESSION AF149783
 VERSION AF149783.1 GI:13897503
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1333)
 AUTHORS Yeh, J.C., Haraoka, N., Petryniak, B., Nakayama, J., Elites, L.G.,
 Rebutka, D., Hindsgraul, O., March, J.D., Lowe, J.B. and Fukuda, M.
 Novel sulfated lymphocyte homing receptors and their control by a
 Core1 extension beta 1,3-N-acetylglucosaminyltransferase
 Cell 105 (7), 957-969 (2001)
 JOURNAL MEDLINE
 PUBMED 21332592
 REFERENCE 2 (bases 1 to 1333)
 AUTHORS Haraoka, N. and Fukuda, M.
 Direct Submission
 Submitted (10-MAY-1999) Glycobiology, The Burnham Institute, 10901
 North Torrey Pines Road, La Jolla, CA 92037, USA
 FEATURES
 source
 location/Qualifiers
 1. 1333
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 93. 1253
 /note="LSSFT"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase"
 /protein_id="AAK48417.1"
 /db_xref="GI:13897504"
 /translation="MLPFRKKMLFLVSOAALALFPHWYSHNLSLMSKAOERH
 VIVLSWRGSSSFVGLFPGHDPFVFLMSPAHMYMPKOSTAMLMAYDILRAFP
 LCDMSVDAYMEPRPROSSIFQWENRALSAPACITIPDELTIPRACHLCSQP
 FEVEKACRSYSHVLEKVERFENLQSLYPLKDPSLNLVHLVPRPAFRVSRERK
 GDLMDIRIYMGQHEOKLREKQDPYVMQYICOSLEIKTIOSLPLALQERYLLVRY
 EDLARAVAOSTRMYEFVGLFPLHLOTVHNITQVGMGSHAFHTPAARDLANTSQAM
 RMSLPYKERVRLQKACGDANMLLGLRYHVRSHQEQRNLLDLSTWYVEQIH"
 BASE COUNT 286 a 393 c 336 g 318 t
 ORIGIN
 Query Match 39.0%; Score 751.2; DB 9; Length 1333;
 Best Local Similarity 76.0%; Pred. No. 1.7e-188;
 Matches 968; Conservative 0; Mismatches 298; Indels 8; Gaps 3;

757 AACAGTCAGCCCTCTTCAGTGGAGCAAGACCGGGCCCTGTGCTCAGCCGCTGTGTG 816
 448 GACAGTCAGCCCTCTTCAGTGGAGCAAGACCGGGCCCTGTGCTCAGCCGCTGTGTG 507
 817 ACTCTTCGCCGCCAGAGATCAGCTCAGCCAGCACTGCAGCTGCTGCGGTGAGC 876
 508 ACATCATCCCAAGATGAATCATCCGCCCGGGCTCAGTGAAGCTCTGTGCGATCAC 567
 877 AGCCCTTGATATGATGAGAGAGAGCCCTGCTCAGCGCTTGTGCTACTCAAGAGG 936
 568 AGCCCTTGATGATGAGAGAGAGCCCTGCTCAGCGCTTGTGCTACTCAAGAGG 627
 937 TGGCTTTTCAGCTGAGCCGCTCATATCCAGTACAGGAGCCCTTCCATCAACCTGC 996
 628 TGGCTTTTCAGCTGAGCCGCTCATATCCAGTACAGGAGCCCTTCCATCAACCTGC 687
 997 AGCTGCTGACCTGCTGAGAGACCCCGGGCCGCTGCTCCAGTCCCGGGAGCAGCACCA 1056
 688 ATATGCTGACCTGCTGAGAGACCCCGGGCCGCTGCTCCAGTCCCGGGAGCAGCAAAAG 747
 1057 TAGAATCATATGATGAGAGATATATGCTGAGGAGCAGTATTTGGAACATCAAGAGG 1116
 748 GAGATCTCATATGATGAGAGATATATGCTGAGGAGCAGTATTTGGAACATCAAGAGG 807
 1117 AAGACAGAGCCCTATTTATGCAATGAGATCATCTGCAAAAGCAGAGTGAGATAGCAAGG 1176
 808 AGGACCAACCTACTATATGATGAGAGATCATCTGCAAAAGCAGTATGATATCAAGAG 867
 1177 CCATCCAAACCTCTGAGAGACCTGTCAGAGCCCTGCTGCTGAGTATGAGAGAC 1236
 868 CCATCCAGTCTTGCAGAGCCCTGTCAGAGACCTGCTGCTGAGTATGAGAGAC 927
 1237 TGGTTCGGGAGACCCCTGCGCCAGAGCAGCAACATATATATTTGTGGGATTTGATTT 1296
 928 TGGTTCGGGAGACCCCTGCGCCAGAGCAGCAACATATATATTTGTGGGATTTGATTT 987
 1297 TGGCCCACTGCAAAAGATGATTTTACATGCTACCCCGGAGAGGAGTGGTGCATG 1356
 988 TGCCCACTGCAAAAGATGATTTTACATGCTACCCCGGAGAGGAGTGGTGCATG 1047
 1357 CTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1416
 1048 CTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1107
 1417 CTTACGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1476
 1108 CTTACGAAAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1167
 1477 ACCTTCAGAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1536
 1168 ACCGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1227
 1537 CCATATCTTGGGAGAGTCTTCCGAGAGATGTAAGAGTGTGTCTGACCCCTTGCTT 1596
 1228 GGACTGTCCTGAGCAAAATCCATAAGAGGTTGTAAGAGGCTTTGCTG--CCACTGTG 1285
 1597 CCAGCTTATGATC 1610
 1286 TCAGCTCATGTCAC 1299

RESULT 9
 AF280088
 LOCUS Homo sapiens L-selectin ligand sulfotransferase GST-3 mRNA, 1992 bp
 DEFINITION complete cds.
 ACCESSION AF280088
 VERSION AF280088.1 GI:12060807
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 1992)
AUTHORS Hemmerich, S., Lee, J.-K., Bhakta, S., Bistrup, A., Ruddle, N.R. and Rosen, S.D.
TITLE Chromosomal localization and genomic organization for the galactose/N-acetylgalactosamine/N-acetylglucosamine 6-O-sulfotransferase gene family
JOURNAL Glycobiology 11 (1), 75-87 (2001)
MEDLINE 21096027
PUBMED 1181564
REFERENCE 2 (bases 1 to 1992)
AUTHORS Hemmerich, S., Lee, J.-K., Bistrup, A., Ruddle, N.R. and Rosen, S.D.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience, 3401 Hillview Ave., Palo Alto, CA 94304, USA
location/Qualifiers
FEATURES
source 1.1992
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q23.1-q23.2"
/cell_type="high endothelial"
/tissue_type="tonsil"
1.1992
/gene="GST3"
122.1282
/gene="GST3"
/note="similar to Homo sapiens sequence encoded by Genbank Accession Number AF131235"
/codon_start=1
/product="L-selectin ligand sulfotransferase GST-3"
/protein_id="AAG48246.1"
/db_xref="GI:12060808"
/translation="MLPKKKLLFLVVSOMALALFPHMSHNSLSMKAPENHVLVLSMRSGSFPVQOLRGQHDVYIMEPANHVMTEFQSTAMLMHVRDLIRVFLDMVDFDIMEGPRKOSLIFOMENSRLCSAPACDILPDEILIPRAKRLCSOOPFEVVKACRSYSHVVLAKVEFENLQSLVPLKDPCLNLIHVLVDRPRAVFRSRETRKGDLMIDSRIVMGHOKLKEKEDPYVMQVVICOSQLEIKTQSLPKALOERYLVRYEDLAPAPVAGTSRMKEFVLEFLPHLQCTVHNITRGKMGDAHFTHNADALNVSOAWRMSLPYEVKSRLOKACGDMNLGLRHVASEQENRLLDLSTMTVPQIH"

BASE COUNT 482 a 540 c 489 g 481 t
ORIGIN

Query Match 38.6%; Score 743.2; DB 9; Length 1992;
Best Local Similarity 76.5%; Pred. No. 2.3e-186;
Matches 938; Conservative 0; Mismatches 283; Indels 5; Gaps 2;

QY 388 TCTCTCTCTCCGAGATGATGCTGTAAAGAAAGGCGTCTGATGTTCTTGGTT 447
DB 105 TCTTCCACTTCAACCAATGCTACTGCTTAAAAAATGAAGCTCTGCTTCTGGTTT 164
QY 448 CCCAGGTCATCGTTAGCTCTCTCATCATATGTCCTCCACAGACACCTTCCAG- 506
DB 165 CCCAGATGGCCATCTTGGCTATTTCTTCACATGACAGCCACCAACATCAGCTCCCTGT 224
QY 507 --AGGAGAGATCCAGAGAGCCCGCTGATGCTGCTGCTCTTCTTGGCGTACGAT 564
DB 225 CTATGAAGACACGCCCGACGCATGACGCTGCTGCTCTTCTTGGCGCTGCGT 284
QY 565 CCTCTTTGTGGAGACGTTTTCGGGACACCGGATGTCTTACCTGATGAGCGCTG 624
DB 285 CTCTCTTTTGTGGGACGCTTTTGGGACACCGGATGTCTTACCTGATGAGCGCG 344
QY 625 CTTGCAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 684
DB 345 CTTGCAATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 404
QY 685 GGGATCTTCTGCTTCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 744
DB 405 GGGATCTGATGACGGGCGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 464

QY 745 CAGCCCCCGGAAACAGTCCAGCTTCTTCAGTGGAGCAAAAGCCGGCCCTGTCTCAG 804
DB 465 CTGGTCCCGGAGACAGTCCAGCTTCTTCAGTGGAGCAAAAGCCGGCCCTGTCTCAG 524
QY 805 CCGCTGTGTGATCTTCTTCCCGCCAGAGATGATGATGATGATGATGATGATGATG 864
DB 525 CACCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 584
QY 865 TCTGCGGTGAGACAGCCCTTGTGATGATGATGATGATGATGATGATGATGATGATG 924
DB 585 TGTGAGTCAAGACGCCCTTGTGATGATGATGATGATGATGATGATGATGATGATG 644
QY 925 TACTCAAGAGAGTGCCTTCTTCTGAGCTGAGGCGCTTATTCATCTACGACGCTT 984
DB 645 TGTCTCAAGAGAGTGCCTTCTTCTGAGCTGAGGCGCTTATTCATCTACGACGCTT 704
QY 985 CCTCAACCTGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1044
DB 705 CCTCAACCTGCAAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 764
QY 1045 AGCAGACCAACATAGAACTCATAGTGTGACAGTGTATGTGCTAGGAGCATTTGGAAA 1104
DB 765 AACGACAAAGAGAGATCTCATGATGATGATGATGATGATGATGATGATGATGATG 824
QY 1105 CGATCAAGAGAGAGACACGCCCTATATGATGATGATGATGATGATGATGATGATG 1164
DB 825 AACTCAAGAGAGAGAGACCAACCTTATGATGATGATGATGATGATGATGATGATG 884
QY 1165 AACTCAAGAGAGAGAGACCAACCTTATGATGATGATGATGATGATGATGATGATG 1224
DB 885 AACTCAAGAGAGAGAGACCAACCTTATGATGATGATGATGATGATGATGATGATG 944
QY 1225 GGTATGAG 1284
DB 945 GGTATGAG 1004
QY 1285 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1344
DB 1005 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1064
QY 1345 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1404
DB 1065 TGGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1124
QY 1405 GTTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1464
DB 1125 GGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1184
QY 1465 ATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1524
DB 1185 ATTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1244
QY 1525 TTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1584
DB 1245 TTTCTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 1304
QY 1585 CACCCCTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1640
DB 1305 --CACCTGTGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTT 1328

RESULT 10
AK026635
LOCUS AK026635 2011 bp mRNA linear PRI 29-SEP-2000
DEFINITION Homo sapiens cDNA: FLJ22982 fls, clone KAT11454, highly similar to AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA.
ACCESSION AK026635
VERSION AK026635.1 GI:10439531
KEYWORDS oligo capping; fls (Full insert sequence).
SOURCE Homo sapiens
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REMARK

Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA
 NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
 Contact: MGC help desk
 Email: cgabs-rr@mail.nih.gov
 Tissue Procurement: Life Technologies, Inc.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
 DNA Sequencing by: Baylor College of Medicine Human Genome Sequencing Center
 Center code: BCM-HGSC
 Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
 Contact: amg@bcm.tmc.edu
 Gunaratne, P.H., Garcia, A.M., Lu, X., Huiyik, S.M., Hale, S.M., Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M., Richards, S., Gibbs, R.A.

FEATURES

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: <http://lmln1.lnl.gov>
 Series: IRM Plate: 50 Row: a Column: 7
 This clone was selected for full length sequencing because it passed the following selection criteria: matched mRNA gi: 5031734.
 Location/Qualifiers
 1. 2037

CDS

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="MGC:34456 IMAGE:5179826"
 /tissue_type="Brain, Lung, Testis, adult, pooled whole"
 /clone_id="NH_MGC_115"
 /lab_host="DH10B"
 /note="vector: pCMV-Sport6"
 126..1238
 /codon_start=1
 /product="Similar to carboxylate (N-acetylglucosamine 6-O) sulfotransferase 4"
 /protein_id="AAH35282.1"
 /db_xref="gi:23273965"
 /translation="MAITLAFPHWTSNHSISLAKAPRHHVLYSSRRSSSTVQ
 LEQHPDVFYLEPAMHWMTEKOSTAMWLAHVRDLRAVELCOMSVDAVMEGR
 ROSSLFOWENRSLASAPACDIIPDEIIPRAHCRLLSQDFEVEKRSYSHVL
 KEVERFNLOSLPLKADPSLNIHVLVDPRAVRSRRTGDMIDRIVMGQHEQ
 KLRKEDOPYVMVQCSOLEYKTLOSIPKALORRYLLVREDIARAPAVATSRYE
 FVGLFELHLOTWNIHNRGKMGDHAFTNRDALNQSANRMLPYKXSRLLKAC
 GDANMLFLHRRHRSQEQRNLLDLISLTYVEQIH"

BASE COUNT

507 a 546 c 461 g 503 t

Query Match 38.6%; Score 743.2; DB 9; Length 2037;
 Best Local Similarity 76.5%; Pred. No. 2.3e-186;
 Matches 938; Conservative 0; Mismatches 283; Indels 5; Gaps 2;

388 TCTCTCTTCGCGAGATGATGCTTGAGAAAGGAGGCTGTGATGTCCTGGGTT 447
 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 61 TTTCCATCTCAGCAAAATGCTACTGCTTAAAAAATGAAGCTCGCTTCTGTTT 120
 448 CCCAGTATCTGTTAGTCTCTCATATGTCCTCCACAGACACCTTCCAG- 506
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 121 CCCAATGCGCATCTGCTCTATCTTCACATGACAGCCACCAACATCAGCTCCTGT 180
 507 --AGGAGAGATCCAGAGAGCCGTCATGCTGCTGCTGCTCTCTCTCTGCGGTGAGAT 564
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 181 CATCAAGCAGACGCCGAGCGCATGACGCTGCTGCTCTCTCTCTGCGGTGAGAT 240
 565 CCTCTTTTGTGGACAGCTTTTTCGGGACAGCACCCGATGCTTCTACATGATGAGCCGTG 624
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 241 CTTCTTTTGTGGGAGAGCTTTTGGGACAGCACCCGATGCTTCTACATGATGAGCCGTG 300
 625 CCTGCGATGTGTGATGATCTTCCACAGAGACAGCTGGAAGTGCATGAGCTGTGTC 684
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 301 CTTGACAGTGTGATGATGATCTTCAAGCAGACGCGCTGATGTCGACAGAGGCTGTGC 360
 685 GGGATCTTCTGCGTCCGCTCTCTGCTGATGATGATGAGCTTTGATGCTCATGAGAAC 744

||||| 111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 361 GGGATCTGATGAGGCGCGCTCTCTCTGTCGACATGAGCTTGTGATGCTCATGAGAAC 420
 745 CAGGCCCCGGGAAGAGTCCAGCGCTTCCAGGGGAGGAAAGCCGGCCGTGCTAG 804
 421 CTGCTCCCGGAGACAGTCCAGCTCTTCTTCAAGGGGAGAGACAGCCGGCCGTGCTAG 480
 805 CCGCTGTGTGATGATCTTCTCCGTCAGAGATGAGTCCAGCAGACATGCAAGCTGC 864
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 481 CACCTGCTGTCATCATCATCCACAAAGATGAATATCCCGCGGCTCACTGAGGCTGC 540
 865 TCTGCGGTGAGAGCCCTTTGATATGTTGAGAGAGGCTGCGCTCTCACGCTTCTG 924
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 541 TGTGAGTCAAGAGCCCTTTGAGTGTGAGAGAGCCGCTGCTCTACAGCAGCTGC 600
 925 TACTCAAGAGAGGCGCTTCTTCAAGCTGACAGCCCTCTATCACTACAGAGACCTT 984
 601 TGTCTAAGAGAGTGGCGCTTCTTCAAGCTGACAGCTCTTACCGCGCTGAGAAAGACCTT 660
 985 CCGTCAACCTGACAGTCTGTCAGCTGTCAGAGACCCCGGCGCTGTCAGATCCCGG 1044
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 661 CCTCAACCTGATATGTCAGCTGTCAGAGACCCCGGCGCTGTCAGATCCCGGAG 720
 1045 AGCAGACCAACATAGACATGATGTTGAGAGATATTTGCTAGGGGACATTTGGAAA 1104
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 721 AACGCAAGAGAGAGATCTCATGATGACAGTCTGATGAGGGGACATGAGACAAA 780
 1105 CGATCAAG 1164
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 781 AACTCAAG 840
 1165 AACTAGTCAAG 1224
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 841 AGATCTACAAAG 900
 1225 GGTATGAG 1284
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 901 GCTATGAG 960
 1285 GGTGAGATTTTGGCCCACTCAAAACATGAGGTTTACATGTCACCCGCGCAAGGCA 1344
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 961 GATTGGAATTTTGGCCCACTCAAAACATGAGGTTTACATGTCACCCGCGCAAGGCA 1020
 1345 TGGGTGAGATGCTTCCATCTACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1404
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 1021 TGGGTGAGATGCTTCCATCTACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1080
 1405 GTTGTGCTTACCTTACGAAAGGTTTCCAGCTTCAAGATGCTGCGGTGAGAGCTATG 1464
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 1081 GCTGTGCTTGGCCCTTACGAAAGGTTTTCGACATTCAGAAAGCTGTGCGATGCCATCA 1140
 1141 ATTGCTGCGGTACCGCCAGCTGATGATGATGATGATGATGATGATGATGATGATG 1200
 1525 TTTGCTGCTTCCATCTATCTTGGGAGAGTCTTCCGAAAGGTTAAGAGGTGTGTCG 1584
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 1201 TTTGCTGCTTCCATCTATCTTGGGAGAGTCTTCCGAAAGGTTAAGAGGTGTGTCG 1260
 1585 CAGCCCTGCTTCAAGCTTATGTCAC 1610
 111 111 111 111 111 111 111 111 111 111 111 111 111 111
 1261 --CCACTGTGTCAAGCTTCAAGTCA 1284

RESULT 12
 BD127258
 LOCUS BD127258 1979 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD127258
 VERSION BD127258.1 GI:23222203
 KEYWORDS JP 200201737-A/2689.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

OY		925	TACGCAAGGAGGTGCCTTTTTCATCGCCGTGGACGGCCCTTAATCCACTCATCAGGACCCTT	984
Dd		592	TGCTCAAGGAGGTGCGCTTTCTTCAACTGCCAGTCCCTCTCAACCCGCTGTAAGAAGCCCTT	651
OY		985	CCCTCAACCTTGCAAGTGTGCACCTGTGTCCGAGACCCCGGGCGGTTCGATCCGGG	1044
OY		1045	AGCACACCACCATGAACATAAGTTATGACAGTAAATTATGTCATGAGGACACATTTGGAAA	1104
Dd		712	AAGCCACAAAGGAGGAGTCTCATGTATGACAATGCATATGTGATGGGACAGATGAGCAA	771
OY		1105	CGATCAAGGAGGAAGACAGACCCCTAATTATGCGCAATGAGATCATCTGCAAAAAAGCCAGGTGG	1164
Dd		772	AACTCAGAAGGAGGAGACCAACCCCTAATGATGATGCAAGTCAATCTGGCAAAAGCAGCTGG	831
OY		1165	ACATATGTCAMGCCATCCAACCTCCCTGTAAGCTGTGCAGACGCGTACTGTTCTGTA	1224
Dd		832	AGATCTCAAGAACCATCCAGTCCCTTGGCCAAAGGCGCTGCAGAAAGCTACTGCTTGTGC	891
OY		1225	GGTATGAGGACCTGTTTCGGGACACCCGTGGGCCAGAGACAGATCATTAATAATTTGTGG	1284
Dd		892	GCTATGAGGACCTTGCTGAGCCCTGTGGCCCCAGACTTCCCAGATATGATTAATTCGTGG	951
OY		1285	GATTGGAATTTTTTGGCCCCACCTCCCAAACTGGGTTTACAATGTCAACCCGGGCAAGGGCA	1344
Dd		952	GATTGGAATTTCTTCCCATCTTCAAGCTGGGTGATCAATCATCAACCCGAGCAAGGGCA	1011
OY		1345	TGGGTCAAGCATGCTTTCATCACTAAGCCAGAGAAAGCGCTCAACAGTCTCTCAGGCGTGGC	1404
Dd		1012	TGGGTGACACAGCTTTCACACAATCCAGAGGATGCCCTTAATGTCTCCAGGCTTGGC	1071
OY		1405	GTTGTGCTTACCTTACGAAAAGGTTTCCGAGCTTCAAGATGCTCGCGGTGAGGCTATGG	1464
Dd		1072	GTGTGTCTTTCCTTATGAAAAGTTTCTGACTTTCAGAAAAGCCTGTGGGATGCCATGA	1131
OY		1465	ATTGTGCTGGATACCTCCAGTAGTCTCAACAGAAACAAGCAGCCTGTCCCTGATC	1524
Dd		1132	ATTGTGCTGGGCTACCGCAGCTGAGATCTGAACAAGAAAGAAACCTGTGCTGTGATC	1191
OY		1525	TTTCTGCTCTCCCTCCATATCTTTGGGAGAGGCTTCCGAAAGGTTAAGAGGCTGTGCTG	1584
Dd		1192	TTTCTGTACTTACCTGGACGTGTCCCTTAAGCAAACTCACTAAGAGGTTGAAAGGCTTGTCTG	1251
OY		1585	CACCCCTTGGTTCCAGCCTTAGTCAC	1610
Dd		1252	--CCACCTGGTGTGACGCTCAGTCAC	1275
RESULT 13				
AK074746				
LOCUS				
DEFINITION	Homo sapiens cDNA FLJ91265 fis, clone NT2RM4001325, weakly similar to CHONDROITIN 6-SULFOTRANSFERASE (EC 2.8.2.17).			
ACCESSION	AK074746.1 GI:22760388			
VERSION	AK074746.1			
KEYWORDS	cDNA; full insert sequence.			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	Isoigai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T., Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S., Kawai,Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y., Kojima,S., Nagahara,K., Masuko,Y., Ono,T., Okeno,K., Yoshikawa,Y., Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and Ninomiya,K.			
TITLE	NEDO human cDNA sequencing project			
JOURNAL	Unpublished			
REFERENCE	2 (bases 1 to 1979)			
AUTHORS	Isoigai,T. and Otsuki,T.			
TITLE	Direct Submission			

JOURNAL

Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute, Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan (E-mail: genomics@hri.co.jp, Tel: 81-438-52-3975, Fax: 81-438-52-3986) NEDO human cDNA sequencing project supported by Ministry of Economy, Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology; cDNA library construction: Institute of Medical Science, University of Tokyo, Laboratory of Gene Structure, Human Genome Center; cDNA 5'- & 3'-end one pass sequencing and clone selection: Helix Research Institute (supported by Japan Key Technology Center etc.).

FEATURES

source

1. 1979
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NT2RM4001325"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/note="cloning vector: pME185FL3-mRNA from uninduced NT2 neuronal precursor cells"
69. 1229
/note="unnamed protein product"
/codon_start=1
/protein_id="BAC11177.1"
/db_xref="GI:22760389"
/translation="MLPKMKLLFLVSOMAILFPMYSHNISLSMKAPERM
LIVLSWRSSEFVQLEQHPDVFYLMPPAHVMTFKQSTAMMLHMAVRDLIRAVF
ICDMSVFAIMERPGRROSLFQWENSRALCSAPACDLIPDEILIPRAVFLRSQOP
FEVNEKACRSYSHVLLKEVFNLOSILYLLKDPISLNIHIVLRPBRFSRETK
GLMDINRIVMGHOKLEKEDOPYVYVICOSELEYKITIOSLPKALOEVLYLVY
EDLARPAVNOTSRMYEYVELEFPLHQLQTYHNTIRKSGMDIAFHINADALNVSOAM
RMSLYEKYSRLQKACGDMNLLGTRHNSBOQRMLDLILSTVTPQIH"

CDS

BASE COUNT 451 a 542 c 474 g 512 t
ORIGIN

Query Match

Best Local Similarity 76.4%; Score 741.6; DB 9; Length 1979;
Matches 937; Conservative 0; Mismatches 284; Indels 5; Gaps 2;

388 TCTCTCTTCCGAGATGATGCTGTGAAGAGGAGGCTGATGATGCTGGGTT 447
111 111 111 111 111 111 111 111 111 111 111 111 111 111 111
52 TCTTCACATTCAGCAATGCTACTGCTTAAAAAATGAAGCTCCTGTTCTGTTT 111
111 111 111 111 111 111 111 111 111 111 111 111 111 111
448 CCCAGGTATCGTTGAGCTCTCTCATATGCTCCGTCACAGACCTTTCCAG- 506
111 111 111 111 111 111 111 111 111 111 111 111 111 111
112 CCCAGATGGCATTCTTGCTATTCTTCCATGATGACGACCAACATCAGTCCCTGT 171
111 111 111 111 111 111 111 111 111 111 111 111 111 111
507 --AGGAGAGATCCAGAGAGCCCGTGCATGTGCTGCTGCTCTGCGGTGAGAT 564
111 111 111 111 111 111 111 111 111 111 111 111 111 111
172 CATTGAAGGACAGCCCGGACGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 231
111 111 111 111 111 111 111 111 111 111 111 111 111 111
565 CCTCTTTTGTGGGACACTTTTGGGACACCCGAGTGTCTTACATGATGAGCCGTG 624
111 111 111 111 111 111 111 111 111 111 111 111 111 111
232 CTTCTTTTGTGGGACACTTTTGGGACACCCGAGTGTCTTACATGATGAGCCGTG 291
111 111 111 111 111 111 111 111 111 111 111 111 111 111
625 CCTGCGATGTGATGATGATTTACAGACAGACACCTGGAAGCTGCATGCTGTGTC 684
111 111 111 111 111 111 111 111 111 111 111 111 111 111
292 CCTGCGACGTGTGATGATGATTTACAGACAGACACCCGAGTGTGTCATGCTGTGTC 351
111 111 111 111 111 111 111 111 111 111 111 111 111 111
685 GGGATCTTCTGCTCCGCTCTCTGATGATGATGATGATGATGATGATGATGATGATG 744
111 111 111 111 111 111 111 111 111 111 111 111 111 111
352 GGGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 411
111 111 111 111 111 111 111 111 111 111 111 111 111 111
745 CAGGCCCCGGAAGACAGTCCAGCTTCTTCCAGTGGAGCAAGCCGCGCTGTGCTCAG 804
111 111 111 111 111 111 111 111 111 111 111 111 111 111
412 CTGGTCCCGGAGACAGTCCAGCTTCTTCCAGTGGAGCAAGCCGCGCTGTGCTCAG 471
111 111 111 111 111 111 111 111 111 111 111 111 111 111
805 CCGCTGTGTGATCTTCTCCCTGCCACAGATCAGCTCACCACAGACATGCAAGCTGC 864
111 111 111 111 111 111 111 111 111 111 111 111 111 111
472 CACCTGCTGTGATCATTATCCCAAGATGAATATCATCCCGGCTCATGCTGCTCC 531
111 111 111 111 111 111 111 111 111 111 111 111 111 111
865 TCTGGGTGAGACAGCCCTTGTATGTGTGAGAGAGCCCTGCGCTCTCAGCGCTTGTG 924
111 111 111 111 111 111 111 111 111 111 111 111 111 111

111 111 111 111 111 111 111 111 111 111 111 111 111 111
532 TGTGCAAGTCAAGAGCCCTTGTGAGTGTGAGAGAGCCCTGCTCTCAGACGAGTGG 591
111 111 111 111 111 111 111 111 111 111 111 111 111 111
925 TACTCAAGAGAGGCTTCTTCCAGCTCAGAGCCCTTATCCACTACACAGACCTT 984
111 111 111 111 111 111 111 111 111 111 111 111 111 111
592 TGTCTCAAGAGAGGCTTCTTCCAGCTCAGAGCCCTTATCCACTACACAGACCTT 651
111 111 111 111 111 111 111 111 111 111 111 111 111 111
985 CCCTCAAGCTGCAAGCTGTCAGCTGTCAGAGCCCGGCGCTGTCCAGTCCGCG 1044
111 111 111 111 111 111 111 111 111 111 111 111 111 111
652 CCCTCAAGCTGCAAGCTGTCAGCTGTCAGAGCCCGGCGCTGTCCAGTCCGCG 711
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1045 AGCACACCAACATAGACTCATGTTGAGAGTATTTGTGTAGGGCAGCATTTGGAAA 1104
111 111 111 111 111 111 111 111 111 111 111 111 111 111
712 AACGCAAGAGAGAGATCATGATGATGATGATGATGATGATGATGATGATGATG 771
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1105 CGATCAAG 1164
111 111 111 111 111 111 111 111 111 111 111 111 111 111
772 AACTCAAG 831
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1165 ACATAGTCAAG 1224
111 111 111 111 111 111 111 111 111 111 111 111 111 111
832 AGATCTACAG 891
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1225 GGTATGAG 1284
111 111 111 111 111 111 111 111 111 111 111 111 111 111
892 GGTATGAG 951
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1285 GGTATGAG 1344
111 111 111 111 111 111 111 111 111 111 111 111 111 111
952 GATGAG 1011
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1345 TGGGTAG 1404
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1012 TGGGTAG 1071
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1405 GGTATGAG 1464
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1072 GGTATGAG 1131
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1465 ATTGCTGAG 1524
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1132 ATTGCTGAG 1191
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1525 TTCTGCT 1584
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1192 TTCTGCT 1251
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1585 CACCCCTGCT 1610
111 111 111 111 111 111 111 111 111 111 111 111 111 111
1252 --CCACTGCT 1275
111 111 111 111 111 111 111 111 111 111 111 111 111 111

RESULT 14
AR203335 2032 bp DNA 1linear PAT 20-JUN-2002
LOCUS AR203335
DEFINITION Sequence 2 from patent US 6365365.
ACCESSION AR203335
VERSION AR203335.1 GI:21499698
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistup,A., Rosen,S.D., Tangemann,K. and Hemmerich,S.
TITLE Method of determining whether an agent modulates glycosyl
sulfotransferase-3
JOURNAL Patent: US 6365365-A 2 02-APR-2002;
FEATURES Location/Qualifiers
source 1..2032
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN

Query Match 38.5%; Score 741.6; DB 6; Length 2032;
 Best Local Similarity 76.4%; Pred. No. 6,3e-186;
 Matches 937; Conservative 0; Mismatches 284; Indels 5; Gaps 2;

```

QY      TCCTCTCTCCGAGATGATGCTGTGAAGAAAGAGAGCTGCTGATGTTCTGGGTT 447
DB      TCTTCCACTTCAGCAATGCTACTGCTTAAAAAAGAAAGAAAGCTGCTGTTGTGTTT 224
QY      448 CCCAGGATGCTGTGATGCTCTTCAATCCATATGCTGCTGACAGACACTTTCCAG- 506
DB      225 CCCAGATGGCCATTTGGCTCTATCTTCTTCCATGTCACAGCCACAAATCAGCTCCCTGT 284
QY      507 --AGGAGAGATCCAGAGAGCCCGTGCATGCTGCTGCTGCTTCTCCGCGGTCAGAT 564
DB      285 CTATGAAGGCACACCCGAGCGCATGCACGCTGCTGCTGCTGCTGCTGCTGCT 344
QY      565 CCTCTTTTGTGGACAGCTTTTGGGAGACACCCGATGTTCTTACCTGATGAGCCCTG 624
DB      345 CTCTCTTTTGTGGGACCTTTTGGGAGACACCCAGATGTTTCTTACCTGATGAGCCCG 404
QY      625 CCTGGCANTGTGATGATGCTTCCACAGCAGCAGACAGCTGGAACCTGACATGGCTGTC 684
DB      405 CCTGGCAGCTGTGATGATGCTTCCACAGCAGCAGCAGCAGCTGATCTGACATGGCTGTC 464
QY      685 GGGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 744
DB      465 GGGATCTGATAGCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 524
QY      745 CAGGCCCCCGGAAACAGTCCAGCTTCTTCCAGTGGAGCAAGCCGGGCTGCTGCTG 804
DB      525 CTGCTCCCGGAGACAGTCCAGCTTCTTCCAGTGGAGCAAGCCGGGCTGCTGCTGCTG 584
QY      805 CGCCTGTGTGACTTCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 864
DB      585 CACCTGCTGTGATGATGCTTCCACAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 644
QY      865 TCTGCGGTGACAGCAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 924
DB      645 TGTGCTGATCAACAGCTTGTGATGATGATGATGATGATGATGATGATGATGATGATG 704
QY      925 TACTCAGAGAGTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 984
DB      705 TGTGCTCAGAGAGTGTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 764
QY      985 CCTCTACAGCTGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1044
DB      765 CCTCTACAGCTGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 824
QY      1045 AGCAGACCCATAGAACTATGCTGATGCTATGCTATGCTATGCTATGCTATGCTATG 1104
DB      825 AACGCAAGAGAGAGATCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 884
QY      1105 CGATCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1164
DB      885 AACGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 944
QY      1165 ACATAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1224
DB      945 AGATCTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1004
QY      1225 GGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1284
DB      1005 GGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1064
QY      1285 GGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1344
DB      1065 GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1124
QY      1345 TGGGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1404
DB      1125 TGGGTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1184
QY      1405 GTTGTCTTACCTTACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1464
  
```

Query Match 38.5%; Score 741.6; DB 6; Length 2032;
 Best Local Similarity 76.4%; Pred. No. 6,3e-186;
 Matches 937; Conservative 0; Mismatches 284; Indels 5; Gaps 2;

```

DB      1185 GCTGTCTTCCGAGATGATGCTGTGAAGAAAGAGAGCTGCTGATGTTCTGGGTT 1244
QY      1465 ATTGCTGGGATTAATCTCCAGTCAAGTCAACAAAGAAAGAGAGAGAGAGAGAGAGAG 1524
DB      1245 ATTGCTGGGATTAATCTCCAGTCAAGTCAACAAAGAAAGAGAGAGAGAGAGAGAG 1304
QY      1525 TTGCTGCTCTCCATATCTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1584
DB      1305 TTGCTGCTCTCCAGTCAAGTCAACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1364
QY      1585 CACCCCTGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1610
DB      1365 --CCACCTGCTCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1388
  
```

RESULT 15
 BD134772
 LOCUS BD134772 2032 bp DNA Linear PAT 18-SEP-2002
 DEFINITION Glycosylsulfoltransferase-3.
 ACCESSION BD134772
 VERSION BD134772.1 GI:23229717
 KEYWORDS JP 2002507409-A/1.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 2032)
 Bistup,A., Rosen,S.D., Tangemann,K. and Hammerich,S.
 Glycosylsulfoltransferase-3.
 Patent: JP 2002507409-A 1 12-MAR-2002;
 JOURNAL THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, SYNTEX INC
 OS Homo sapiens (human)
 PN JP 2002507409-A/1
 PD 12-MAR-2002
 PR 26-FEB-1999 JP 2000537979
 PF 20-MAR-1998 US 09/045284,12-NOV-1998 US 09/190911 PI
 ANNETTE BISTRUP, STEVEN D ROSEN, KIRSTEN TANGEMANN, STEFAN PI
 HAMMERICH
 PC C12N9/10,A01K67/027,A61K45/00,A61K48/00,A61P29/00,A61P37/06,
 PC A61P43/00,
 PC C12N1/15,C12N1/21,C12N5/10,C12N15/09,C12Q1/48,C12N5/00,C12N15/
 PC 00
 CC Glycosylsulfoltransferase-3
 FH Key
 FT source 1.2032
 FT Location/Qualifiers
 FT Location/Qualifiers
 FT 1.2032
 FT /organism="Homo sapiens"
 FT /mol_type="genomic DNA"
 FT /db_xref="taxon:9606"

BASE COUNT 468 a 569 c 490 g 505 t

ORIGIN

Query Match 38.5%; Score 741.6; DB 6; Length 2032;
 Best Local Similarity 76.4%; Pred. No. 6,3e-186;
 Matches 937; Conservative 0; Mismatches 284; Indels 5; Gaps 2;

```

QY      388 TCTCTCTTCCGAGATGATGCTGTGAAGAAAGAGAGCTGCTGATGTTCTGGGTT 447
DB      TCTTCCACTTCAGCAATGCTACTGCTTAAAAAAGAAAGAAAGCTGCTGTTGTGTTT 224
QY      448 CCCAGGATGCTGTGATGCTCTTCAATCCATATGCTGCTGACAGACACTTTCCAG- 506
DB      225 CCCAGATGGCCATTTGGCTCTATCTTCTTCCATGTCACAGCCACAAATCAGCTCCCTGT 284
QY      507 --AGGAGAGATCCAGAGAGCCCGTGCATGCTGCTGCTGCTTCTCCGCGGTCAGAT 564
DB      285 CTATGAAGGCACACCCGAGCGCATGCACGCTGCTGCTGCTGCTGCTGCTGCTGCT 344
QY      565 CCTCTTTTGTGGACAGCTTTTGGGAGACACCCGATGTTCTTACCTGATGAGCCCTG 624
  
```

Db 345 CTTCTTTTGTGGGGCAGCTTTTGGGAGACCCAGATGTTTCTACTGATGAGCCCG 404
QY 625 CCTGCATGTGTGGATGACTTTGCACGACAGACCTGGAACTGCACATGGCTGTGC 684
Db 405 CCTGGCAGCTGTGGATGACTTTGCACGACAGACCTGGATGCTGCATGAGCTGTGC 464
QY 685 GGGATCTTCTGCTTCCGCTTCTCTGTGACATGAGCGCTTGTGATGCTTACATGACC 744
Db 465 GGGATCTGATGAGCGGCTTCTCTGTGACATGAGCGCTTGTGATGCTTACATGACC 524
QY 745 CAGGCCCCGGAAACAGTCCAGCTTCCAGTGGAGCAAAAGCCGGCCCTGTGCTAG 804
Db 525 CTGTGCCCCGGAGACAGTCCAGCTTCTTCAAGTGGAGAAACAGCCGGCCCTGTGCTG 584
QY 805 CCGCTGTGTGTGACTTCTTCCCTGCCCAGAGATCAGCTCACCAGCACTGCAAGCTGC 864
Db 585 CACCTGCTGTGACATCATCCCAAGATGAATCATCCCGCGGCTCACTGACAGCTCC 644
QY 865 TCTGCGGTGAGCAGCCCTTGTGATGATGAGAGAGCGCTGCTGCAAGCGCTGTG 924
Db 645 TGTGCACTCAACAGCCCTTGTGAGGTGTGAGAGAGCGCTGCGCTCTCTACAGCAGCTG 704
QY 925 TACTCAAGAGGTGCTTTTCTGAGCTGAGCGCCCTATCCACTACTCAGCAGCCCT 984
Db 705 TGTCTCAAGAGGTGCGCTTCTTCAACCTGAGTCCCTTACCGCTGCTAAAGACCCCT 764
QY 985 CCTCAAGCTGCAAGCTGCTGACCTGCTGAGAGACCCCGGCGCTGTCCGATCCCGGG 1044
Db 765 CCTCAAGCTGCAAGCTGCTGACCTGCTGAGAGACCCCGGCGCTGTCCGATCCCGAG 824
QY 1045 AGCAGACCAACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1104
Db 825 AACGACAAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 884
QY 1105 CGATCAAG 1164
Db 885 AACTCAAG 944
QY 1165 ACATGATCAAG 1224
Db 945 AGATCTCAAG 1004
QY 1225 GGTATGAG 1284
Db 1005 GCTATGAG 1064
QY 1285 GGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1344
Db 1065 GATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1124
QY 1345 TGGGTGAG 1404
Db 1125 TGGGTGAG 1184
QY 1405 GTTGTGCTTACCTTACGAAAGGTGTTCCAGCTTCAAGATGCTGCGGTGAGAGTATGG 1464
Db 1185 GCTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCTTGTGCT 1244
QY 1465 ATTTGCTGGAGATCACTCAGTCAAGATCACTCAGTCAAGATCACTCAGTCAAGATCA 1524
Db 1245 ATTTGCTGGAGATCACTCAGTCAAGATCACTCAGTCAAGATCACTCAGTCAAGATCA 1304
QY 1525 TTTCTGCTCTCTCCATCTTGTGGGAGAGTCTTCCGAGAGGTTAAGAGAGTCTGCTG 1584
Db 1305 TTTCTGCTCTCTCCATCTTGTGGGAGAGTCTTCCGAGAGGTTAAGAGAGTCTGCTG 1364
QY 1585 CACCCCTTGTGCTCAGCTTATGCTAC 1610
Db 1365 --CCACTGTGTGCTCAGCTTATGCTAC 1388

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 10:39:13 : Search time 478.467 Seconds

(without alignments)
10866.196 Million cell updates/sec

Title: US-09-645-078-3
Perfect score: 1926
Sequence: 1 gggcactcaactacactg.....ccgtcactcatttttaag 1926

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*

- 1: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1980.DAT:*
- 2: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1981.DAT:*
- 3: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1982.DAT:*
- 4: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1983.DAT:*
- 5: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1984.DAT:*
- 6: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1985.DAT:*
- 7: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1986.DAT:*
- 8: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1987.DAT:*
- 9: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1988.DAT:*
- 10: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1989.DAT:*
- 11: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1990.DAT:*
- 12: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1991.DAT:*
- 13: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1992.DAT:*
- 14: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1993.DAT:*
- 15: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1994.DAT:*
- 16: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1995.DAT:*
- 17: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1996.DAT:*
- 18: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1997.DAT:*
- 19: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1998.DAT:*
- 20: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA1999.DAT:*
- 21: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2000.DAT:*
- 22: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001A.DAT:*
- 23: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2001B.DAT:*
- 24: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2002.DAT:*
- 25: /SIDSI/gcgdata/geneseq/geneseq-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1926	100.0	1926	20	AA220793
2	751.2	39.0	1333	24	AA516947
3	741.6	38.5	1979	22	AAK94229
4	741.6	38.5	2032	20	AA220792
5	717	37.2	2065	21	AA294211
6	468.8	24.3	877	22	AAK91803
7	468.8	24.3	877	22	AAK93921
8	429.2	22.3	2988	21	AAK76156

9	370.6	19.2	1647	24	AA224670
10	370.6	19.2	1694	22	AA202700
11	370.6	19.2	2544	24	ABN89506
12	370.6	19.2	160552	22	AA202697
13	355.6	18.5	1937	22	AA516948
14	355.6	18.5	1989	22	AA202696
15	354	18.4	2044	22	AA202699
16	354	18.4	2170	22	AA202698
17	304.2	15.8	48436	24	ABN89533
18	118.6	6.2	668	24	ABQ21506
19	118.6	6.2	668	24	ABQ21507
20	114	5.9	1458	19	AAV36418
21	114	5.9	2190	24	AAV41280
22	114	5.9	2415	25	ABX08787
23	110.6	5.7	7099	25	ACC46083
24	107.4	5.6	2156	19	AAV21200
25	98.4	5.1	2354	18	AAK5037
26	96.2	5.0	524	24	ABQ21510
27	96.2	5.0	524	24	ABQ21511
28	95.8	5.0	3029	22	AAH17932
29	95.4	5.0	524	24	ABQ21512
30	95.4	5.0	524	24	ABQ21513
31	94	4.9	2393	24	AA561699
32	94	4.9	2409	20	AAK87821
33	88.6	4.6	2150	20	AAK87820
34	86.8	4.5	668	24	ABQ21508
35	86.8	4.5	668	24	ABQ21509
36	82	4.3	2165	25	AB221794
37	79.8	4.1	576	24	ABQ45408
38	79.8	4.1	576	24	ABQ45409
39	65	3.4	65	24	ABN54495
40	63.2	3.3	576	24	ABN54496
41	63.2	3.3	576	24	ABN54497
42	52.4	2.7	731	20	AA224563
43	52.4	2.7	731	21	AAK5802
44	52.4	2.7	731	24	ABQ92207
45	52.4	2.7	731	24	ABL49021

ALIGNMENTS

RESULT 1	
AA220793	
ID	AA220793 standard; DNA; 1926 BP.
AC	AA220793;
DT	08-DEC-1999 (first entry)
XX	
XX	Mouse glycosyl sulfotransferase-3 coding sequence.
DE	
XX	
XX	Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
KW	selectin binding interaction; inflammation; lymphocyte homing; mouse;
KW	secondary lymph organ; ss.
XX	
OS	Mus sp.
XX	
XX	W09949018-A1.
XX	
XX	30-SEP-1999.
XX	
XX	26-FEB-1999; 99WO-US04316.
XX	
XX	20-MAR-1998; 98US-0045284.
PR	12-NOV-1998; 98US-0190911.
XX	
PA	(REGC) UNIV CALIFORNIA.
XX	(SYNT) SYNTAX USA INC.
XX	
PI	Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
XX	
DR	WPI; 1999-580442/49.

Oy	1961	TACGCGACACGCAAAACAGCATCTGCTGAGCTGCTGTAAACCTCCCTGCACATCTTT	1920
Db	1861	TACTGGACACAGCAAAACAGCATCTGTGTGAGTGTGTCTGTAAACCTCCCTGTACATCTTT	1920
Oy	1921	TCTAAG	1926
Db	1921	TCTAAG	1926
RESULT 2			
ID	AA516947		
XX	AA516947	standard; cDNA; 1333 BP.	
AC	XX		
XX	AA516947;		
DT	XX		
XX	12-MAR-2002	(first entry)	
DE	XX		
XX	Human L-selectin sulfotransferase-2 (LSS2-2) cDNA.		
KW	Human; beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;		
KW	L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;		
KW	ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;		
KW	allergic contact dermatitis; lymphoma; chronic pneumonia; LSS2-2;		
KW	delayed-type hypersensitivity reaction; hyperplastic thymus; antileuc;		
KW	antiinflammatory; antipsoriatic; antidiabetic; dermatological;		
KW	antiallergic.		
XX			
OS	Homo sapiens.		
XX			
FH	Key	Location/Qualifiers	
FT	CDS	111..1253	
FT		/*tag= a	
FT		/product= "Human LSS2-2"	
XX			
PN	WO200185177-A1.		
XX			
PD	15-NOV-2001.		
XX			
PF	10-MAY-2001; 2001WO-US15452.		
XX			
PR	11-MAY-2000; 2000US-0569320.		
XX			
PA	(BURN-) BURNHAM INST.		
XX			
PI	Fukuda M, Yeh J, Hiraoka N;		
XX			
DR	WPI; 2002-075226/10.		
XX	P-PSDB; AAU11274.		
PT	New enzyme, useful for modifying acceptor molecule, comprises an		
PT	isolated L-selectin sulfotransferase-2 that directs expression of		
PT	L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or		
XX	intestinal G1cM6C 6-sulfotransferase		
PS	Claim 19; Fig 4; 98pp; English.		
XX			
CC	The present invention provides a method of modifying an acceptor molecule		
CC	by contacting the acceptor with an isolated		
CC	beta1,3-N-acetylglucosaminyltransferase (beta1,3GNT) or an active		
CC	fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The		
CC	invention also provides a method of treating or preventing an		
CC	L-selectin-mediated condition by reducing the expression or activity of a		
CC	beta1,3GNT that directs expression of a MECA-79 antigen. This can be done		
CC	by administering to the subject an oligosaccharide L-selectin antagonist		
CC	that inhibits binding of L-selectin to a MECA-79 antigen, for example by		
CC	administering antibody material that specifically binds beta1,3GNT,		
CC	and/or a beta1,3GNT antisense nucleic acid molecule. L-selectin		
CC	sulfotransferase-2 (LSS2-2) also directs MECA-79 antigen expression.		
CC	Alternatively, the expression or activity of LSS2-2 or its active		
CC	fragment can be reduced in combination with reducing the expression or		
CC	activity of beta1,3GNT. The method is useful for treating L-selectin		
CC	mediated conditions such as Crohn's disease and ulcerative colitis,		
CC	inflammatory disorders of the skin such as allergic contact dermatitis,		

CC	psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type
CC	hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC	sequence represents cDNA encoding human LST-2.
XX	
SQ	Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;

Query Match	39.0%;	Score 751.2;	DB 24;	Length 1333;
Best Local Similarity	76.0%;	Pred. No. 6.5e-206;		
Matches 968;	Conservative	0;	Mismatches 298;	Indels 8; Gaps 3
OY	340 TAACAACTAAGTACTGAGCCCCATCCCCTGTGGTCCTTTCAAGGCTCTTCTCCTTCC	399		
Db	31 TAAAACAGACGCCAACATCCACCCTTTCTGTGTTCTCTTAAGGCTCTT---CACTTCA	87		
OY	400 GCAGAGTATGCTGTTGAAGAAAGGAGGAGCTGCATGATTCTCTGAGTTCCAGTGATCG	459		
Db	88 GCACAATGCTACTGCTTAAAAAATAAGAGCTCTCTGTCTTCTGTCCAGATAGGCCA	147		
OY	460 TTGTAGCTCTCTCATATGTCCGTCACAGACACTTCTCCAG---AGGAGAGGT	516		
Db	148 TCTTGGCTCTATCTTCCACATGTACAGCCACACATACAGCTCCCTGTCTATGAAGGCAC	207		
OY	517 CCAGAGAGCCCGTAGTGTGTGTGTGTCTTCCCTGGGCGTCAGAGATCTTCTTTGTGG	576		
Db	208 AGCCGAGAGCGATGACAGCGTGTGTGTGTCTCTCTGGGCGCTGAGCTTCTTTGTGG	267		
OY	577 GACAGCTTTTGGGCGACAGACCCGGATGTGTTCTTACCTGATGAGAGCTTGCCTGGATGTGT	636		
Db	268 GGCACTCTTTTGGGCGACACCCAGATGTTTTCTTACCTGATGAGAGCCCGCTGGCACGTGT	327		
OY	637 GGATGACTTTTACCGACAGACAGAGCTGGAACTGACATAGGCTGTGGGGATCTTCTGC	696		
Db	328 GGATGACTTTTACCGACAGACAGAGCTGGAACTGACATAGGCTGTGGGGATCTTCTGC	387		
OY	697 GTTCCGCTTCCCTGTGTGACATGAGAGCTTCTTGTATGCTTACATGAAACCCAGGCCCCGGA	756		
Db	388 GGGCGGCTCTTCTGTGTGACATGAGAGCTTGTGATGCTTACATGGAACCTGGTCCCGGA	447		
OY	757 AACAGTCCAGAGCTTTCCAGTGGGAGCAAAGCCGGGCGCTGTGCTACAGGCTGTGTGTG	816		
Db	448 GACATCTCAGAGCTTTTTCAGTGGGAGAAAGCCGGGCGCTGTGTTCTGACAGCTGCTGTG	507		
OY	817 ACTTCTTCCCTGCCACGAGATCAGCTTACCCACAGCATGGAAGGTGCTGTGGGTCAGC	876		
Db	508 ACATATCTCCACAGATGAATATATCCCGGGGCTACTGACAGGCTCTGTGACATGAC	567		
OY	877 AGCCCTTGTATATGTGTGAGAAGGCTGCCGCTTACAGGCTTCTGTGATCTCAAGAGAG	936		
Db	568 AGCCCTTGTATATGTGTGAGAAGGCTGCCGCTTCTACAGGCTGTGTGATCTCAAGAGAG	627		
OY	937 TGCGTTTCTCAGAGCTGAGGCGCTTATCAGTACTACAGGAGCCCTTCCCTCAACCTGC	996		
Db	628 TGCGTTTCTCAGAGCTGAGGCTTATCAGTACTACAGGAGCCCTTCCCTCAACCTGC	687		
OY	997 AGCTGTGACCTGTGTGAGAGACCCCGGGCGCTGTCCGATCCCGGAGACACACCA	1055		
Db	688 AATGTGTGACCTGTGTGAGAGACCCCGGGCGCTGTCCGATCCCGGAGACACCAAAAG	747		
OY	1057 TGAAGCTATAGGTTGACAGTCAATATGTTGCTAGGAGAGCATTTGGAAAGATCAAGAGAG	1111		
Db	748 GAGATCTCATGATGACAGTCCCATATGTGTATGAGGAGCATAGCAAAAACATCAAGAAAG	807		
OY	1117 AAGACAGAGCCCTTATGCTATGAAAGATATCTGCAAAAGCCAGGTGACATAGTCAAG	1177		
Db	808 AAGACCAACCTACATATGATGAGAGTATCTGGCAAAAGCAGCTGGAGATCTCAAGA	867		
OY	1177 CCATCCAACCTTCCCTGAGCTTGCAGACAGGCTTACTGTTCTGAGGTATGAGAGAC	1233		
Db	868 CCATCCAAGCTTGGCCCAAGGCGCTGACAGAGAGCTTACTGTTGCTGTATGAGAGAC	927		
OY	1237 TGCTTGGGACACCCCTGGGCGCAGACAGACCAAGCATATTAATTTGGGTTGGGATTTT	1297		
Db	928 TGCTTGGGACCCCTGTGGCCACAGACTTCCCGAATGATATAATTCGTGGGATGGAATTC	987		

QY 1297 TGCCCCACCTCCAAACATGGTTTACATGTCACCCGCGCAAGGGCATGGTCACCATG 1356
 DB 988 TGCCCATCTTCAGACCTGGTGACATACATCACCGAGGCAAGGGCATGGTGACACAG 1047
 QY 1357 CCTTCATCTACTAAGCCGAGGAAGCCCTCAACGCTCTCTAGGCGGTGGCTCTTAC 1416
 DB 1048 CTTTCCACCAATATGCGAGGATGCGCTTAATGTCTCCAGAGGCTGGCTCTTTC 1107
 QY 1417 CTACGAAAGGTTTCCAGCTTCAAGATGCTGGCGGTATGGATGTTGCTGGAT 1476
 DB 1108 CATTGAAAAGTTTCTCAGCTTCAAGAAAGCCGTGGCATGCAATTTGCTGGCT 1167
 QY 1477 ACCCTCAGGTCAAGTCTCAACAGAAAGCAAGCAAGCAAGCAAGCAAGCAAGCAAG 1536
 DB 1168 ACCGCAAGCTCAGATCTGAACAAGAACAGAAAGCAAGCAAGCAAGCAAGCAAGCAAG 1227
 QY 1537 CCCATATCTTGGGGGAGCTTCCGAGGAGTTAAGAGGTCTGTGCAACCCCTTGAT 1596
 DB 1228 GAGCTGCTCCCTGAGCAATCTCAAGAGGTTGAGAGGCTTGTCTG--CCACCTGTG 1285
 QY 1597 CCAGCTTAGTCAC 1610
 DB 1286 TCAGCCTCAGTCAC 1299

RESULT 3
 AAK94229
 ID AAK94229 standard; cDNA; 1979 BP.

XX AAK94229;

DT 06-NOV-2001 (first entry)

XX Human full-length cDNA, SEQ ID NO: 2816.

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX Homo sapiens.

XX EP130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

XX 11-JAN-2000; 2000JP-0118774.

XX 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;

XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX MPI: 2001-524255/58.

XX P-PSDB; AAK93309.

XX 830 Primers useful for synthesizing full length cDNA clones and their

XX use in genetic manipulation -

XX Claim 8: SEQ ID NO 2816; 1380bp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA

XX clones. 830 cDNA molecules encoding a human protein have been

XX isolated and nucleotide sequences of 5' - and 3' - ends of the cDNA

XX molecules have been determined. Primers for synthesizing the full length

XX cDNA are useful for clarifying the function of the protein encoded by

XX the cDNA. The full length clones were obtained by construction of full

XX length enriched cDNA libraries that were synthesized by the oligo-capping

XX method. The primers enable the production of the full length cDNA easily

XX without any special methods. The present sequence is a full length

XX human cDNA of the invention.

CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX

SO Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other;

Query Match 38.5%; Score 741.6; DB 22; Length 1979;

Best Local Similarity 76.4%; Pred. No. 4,8e-203;

Matches 937; Conservative 0; Mismatches 284; Indels 5; Gaps 2;

QY 388 TCTCCCTTCCGACATGATGCTGTTGAAGAAAGGAGGCGCTGATGCTGGT 447
 DB 52 TCTCCACTCAGCAATGCTAGCTCTAATAAATAAGAGCTGCTGTTGTT 111
 QY 448 CCCAGCTATCGTTGTAGCTCTTTCATCATATGTCGTCACAGACACTTCCAG- 506
 DB 112 CCGAGATGCGCATCTTGCTCTATCTTCCACATGTAACGCAACATCAGCTCCTG 171
 QY 507 --AGGAGAGTCCAGAGAGCCCGTCATGCTGCTGCTCTCTGCGGTCAAGAT 564
 DB 172 CTATGAAGGACACAGCCCGAGCGCATCAGTGTGCTTCTGCTCTGCGCTG 231
 QY 565 CCTCTTTGGGAGACCTTTTGGGACAGACCGGATGATGTTCTACATGATGAGCCTG 624
 DB 232 CTTCTTTTGGGAGACCTTTTGGGACAGACCGGATGATGTTCTACATGATGAGCCTG 291
 QY 625 CCTGGCATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 684
 DB 292 CCGGACGCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 351
 QY 685 GGGATCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 744
 DB 352 GGGATCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 411
 QY 745 CAGGCCCCGGAAGACAGTCCAGGCTTTCAGTGGAGAGAAAGCCGCGCTGCTGAG 804
 DB 412 CTGTGCTCCGAGAGACAGTCCAGGCTTTCAGTGGAGAGAAAGCCGCGCTGCTGAG 471
 QY 805 CGCCTGTGTGACTTCTTCCCTGAGAGAGATCAGTCCAGGCTGCTGAGAGAGAGAGAGAG 864
 DB 472 CAGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 531
 QY 865 TCTGGGTCAGAGCCCTTGTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 924
 DB 532 TGTGAGTCAACAGCCCTTGTAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAGTGTGAG 591
 QY 925 TACTCAAGAGGTGCTTCTTCTCAGCTGAGGCTTATCTATCTATCTATCTATCTATCTAT 984
 DB 592 TGTCTCAAGAGAGGTGCTTCTTCTCAGCTGAGGCTTATCTATCTATCTATCTATCTATCT 651
 QY 985 CCTCAAGCTGACAGTGTGACAGTGTGACAGTGTGACAGTGTGACAGTGTGACAGTGTGAC 1044
 DB 652 CCTCAAGCTGACAGTGTGACAGTGTGACAGTGTGACAGTGTGACAGTGTGACAGTGTGAC 711
 QY 1045 AGCAGACCCAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1104
 DB 712 AACGCAAAAGGAGAGTCTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 771
 QY 1105 CGATCAAG 1164
 DB 772 AACTCAAG 831
 QY 1165 ACATAGTCAAG 1224
 DB 832 AGATCTACAG 891
 QY 1225 GGTATGAG 1284
 DB 892 GCTATGAG 951
 QY 1285 GGTATGAG 1344
 DB 952 GATGAG 1011

QY	1465	ATTTCGCGGGGTTACTCTCCAGGTCAGATCTCAACAAAGAACGGCAACCTGTCCTCGATC	152
Db	1245	ATTTCGCGGGGTTACTCTCCAGGTCAGATCTCAACAAAGAACGGCAACCTGTCCTCGATC	1304
QY	1525	TTCTGTCCTCCCTCCCATATCTTGGGGCAGGTCTTCCGAGAAAGTTAAGAGGTCTGCTG	1584
Db	1305	TTCTGTCCTCCCTCCCATATCTTGGGGCAGGTCTTCCGAGAAAGTTAAGAGGTCTGCTG	1364
QY	1585	CACCCCTTGGTCTCAGGCTTAGTCAAC	1610
Db	1365	--CCACCTGGTGTGAGCCTCAGTCAAC	1388
RESULT 5			
ID	AA294211	standard; cDNA; 2065 BP.	
XX	AA294211;		
DT	19-JUN-2000	(first entry)	
XX	Human transferase TRNSFS-11 cDNA clone 2617407CBI.		
XX	transferase; TRNSFS-11; human; antitumor; cell proliferation;		
KW	inflammation; gastrointestinal disorder; developmental disorder;		
KW	genetic disorder; neurological disorder; reproductive disorder;		
KW	smooth muscle disorder; immunological disorder; gene therapy;		
KW	diagnosis; N-acetylglyucosamine 6-O-sulfotransferase; ss.		
XX			
OS	Homo sapiens.		
XX			
XX	Key	Location/Qualifiers	
FT	CDS	174..1334	
FT		/*tag= a	
XX			
PN	WO200014251-A2.		
PD	16-MAR-2000.		
XX			
PF	09-SEP-1999;	99WO-US209989.	
PR	10-SEP-1998;	98US-0150657.	
PR	04-NOV-1998;	98US-0186779.	
PR	11-MAY-1999;	99US-0133642.	
XX			
XX	(INCY-) INCYTE PHARM INC.		
PI	Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;		
PI	Hallman JL, Azimzal Y;		
DR	WPI; 2000-256996/22.		
DR	P-Psdb; AAY79219.		
PT	Human transferase proteins useful for preventing, diagnosing and		
PT	treating cancers and developmental, gastrointestinal, genetic, and		
PT	immunological, neurological, reproductive and smooth muscle disorders -		
XX	Claim 9; Page 104-105; 113pp; English.		
XX	The present sequence is that of cDNA clone 2617407CBI encoding		
CC	human transferase TRNSFS-11 (see AAY79219), 1 of 15 claimed human		
CC	transferase proteins of the invention (see AAY79209-23). The clone		
CC	was isolated from gall bladder cDNA library GBANOT01. TRNSFS-11		
CC	is expressed in dermatologic and gastrointestinal tissues,		
CC	especially those associated with inflammation and cell		
CC	proliferation. It shows homology to mouse N-acetylglyucosamine		
CC	6-O-sulfotransferase. A polynucleotide comprising nucleotides		
CC	264-333 or 1272-1331 of the present sequence can be used as a DNA		
CC	probe. The new human transferases and polynucleotides can be used		
CC	in the diagnosis, prevention and treatment (including gene therapy		
CC	and antisense therapy) of cancer, developmental disorders,		
CC	gastrointestinal disorders, genetic disorders, immunological		
CC	disorders, neurological disorders, reproductive disorders, and		

XX	Sequence	2065 BP: 466 A; 575 C; 491 G; 533 T; 0 other:	smooth muscle disorders.
SO	Query Match Best Local Similarity Matches 935; Conservative 0; Pmid. No. 6.2e-196; 76.2%; 0; Mismatches 285; Indels 7; Gaps 4;		
OY	388 TCTCTCTTCTCCGAGATGATGCTGTGTGAAGAAAGGAGGCTGTGATGTTCTGGGTT	37.2%; Score 717; DB 21; Length 2065;	447
Db	157 TCTTCCACCTTCAGACATGCTACTGCTCAAAAAAATMACTCCTGCTGTTCTGGTTT		216
OY	448 CCCAGTCATGCTGTGATGCTCTCTTCATCCATATGTCGTGCACAGACACCTTCCAG -		506
Db	217 CCCAGATGAGCCATCTTGCTCTTATTTCTCCACATGTACAGCCACAACTACGCTCCGT		276
OY	507 - -AGGAGAGATCCAGAGAGGCCCGTGCATGTGCTGTGCTCTCTCTCGGCGTCAAGAT		564
Db	277 CTATGAAGGACAGCCGAGCGCATGCACGTGTGTGTTCTGTCTTCTGCGGCTGTGCT		336
OY	565 CCTCTTTTGTGGACAGCTTTTGGGACAGCACCCGGATGTGTTTACTGATGGACCTG		624
Db	337 CTCTTTTGTGGGACAGCTTTTGGGACAGCACCCAGATGTTTCTACTGTATGGAGCCG		396
OY	625 CCTGAGCATGTGTGATGCTTTCACACACACAGACAGCCTGGAGGTGCATAGGCTGTGC		684
Db	397 CCTGAGCATGTGTGATGCTTTCACACACACAGACAGCCTGGATGTGTGCATAGGCTGTGC		456
OY	685 GGGATCTTTCGCTGCTCTCTCTGTGTGACATGAGCGTCTTGTATGCTCATGAAC		744
Db	457 GGGATCTTTCGCTGCTCTCTCTGTGTGACATGAGCGTCTTGTATGCTCATGAAC		516
OY	745 CAGGCCCCCGGAAACAGTCCAGCCTCTCCAGTGGAGCAAGCCGGGCCCTGTGTCTAG		804
Db	517 CTGTGTCCCGGAGACAGTCCAGCCTCTCTCTGTGTGAGAAACAGCCGGGCCCTGTGTCTG		576
OY	805 CGCCTGTGTGTGATCTTCTTCCCTGCCACAGATACAGTTCACCCAAAGCATGCAAGCTGC		864
Db	577 CACCTGCTGTGATCATCATCCACAAAGATGA - ATCATCCCGGGGCTACATGCAAGCTGC		635
OY	865 TCTGCGGTACAGACCCCTTTG - ATATGTGTGAGAAAGCCCTCCGCTTCACAGGCTTCGTG		923
Db	636 TGTGACAGTCAACAGCCCTTTGAAGTGTGTGAGAAAGCCCTCCGCTTCACAGCAAGTGTG		695
OY	924 GTACTCAAGAGGTGCTTTTCTTCAAGCTGCAAGCCCTCTATCACTACTACAGGACCT		983
Db	696 GTGCTCAAGAGGTGCTGCTTCTTCAACCTGCAAGTCCCTCTACCCGCTGTCAAAAGACCCC		755
OY	984 TCCCTCAACCGACAGTGTGCTGACCTGGTCCGAGACCCCGGGGCGTTCGATCCGCGG		1043
Db	756 TCCCTCAACCGACATGTGTGACCTGGTCCGAGACCCCGGGGCGTTCGATCCGCGG		815
OY	1044 GAGCACAACACATAGAAATCATGTTGACAGTATATGTCTTAGGCGACATTTGGAA		1103
Db	816 GAAAGCAACAAAGGAGATCTATGATTAAGTGTGATGTGATGGGACACATGAGAGG		875
OY	1104 ACGATCAAGAGAGAGAACAGCCCTATTTATGCCATGAAGATCATCTGCAAAAGCCAGGTG		1163
Db	876 AAATCTCAAGAGAGAGAACAGCCCTACTATGTATGTATGACGTATCTGCCAAAGCCAGCTG		935
OY	1164 GACATAGTCAAGGCGCATTCAAACCTCTCCCTGAAGCTGTGACAGAGCGGTACTGTCTGTG		1223
Db	936 GAGATCTCAAGAGAGACATTCACATCTTGTGCCCAAGGCCCTGTGACAGAGCGTACTCTGTGTG		995
OY	1224 AGGATATGAGGACCTGTGTGGGACACCCCTGGCCAGAGACACACATATTAATTTGTG		1283
Db	996 CGTATATGAGGACCTGTGTGGGACACCCCTGTGGCCAGACATTCGCAATGTATGAATTCGTG		1055
OY	1284 GGGTGTGATTTTGTGCCCCACCTTCAAAACATGTGGGTTTACATGTACACCCGCGGCAAGGC		1343
Db	1056 GGATTTGGAATTTCTTGCCCACTTTCAGACCTGGGTGATTAACATCAACCCGAGGCAAGGCC		1145
OY	1344 ATGGGTACGACGCTTTCATACTAACAAGCCAGAACGCCCTCAAGCTCTTCAGCGGTG		1403

824 CCCTGCCACGAGATCAGCTCACCACAAGCACTGCAAGCTGCTCTGCGGTCAGCAGCCCT

XX

XX

1124 GCCCTATTATGCCATGAGATCATCTGCCAAAGCCAGGTGGACATAGTCAAGGCATCCA

FT exon 46634..46700

FR1 Aug 15 08:09:40 2003

us-09-645-078-3.rng

Page 17

DB 1178 CATACTTGCGTAAAGAAATGGCGGCAAGCTCTCCAGAGCGCTGGCGCCCAAGCGTTCCCTCTG 1237

QY 1422 GAAAAGGTTTCCCACTTCACAGATGTCGCCGCTGATGAGTATGATGATTTCTGGGATACCTC 1481

DB 1238 ACTAAGATCCCTGCGCGCTCTCAGAGATGTGCGCCGCGCGCTCAGCTCTCTGGCTACCGG 1297

QY 1482 CAGGTCCAGATTCCTCAACAGAACAAAGCGCAACTGTCTCCCTGGATCTTTCTG 1529

DB 1298 CCTGTGTACTCTGGCGGACACAGCGATGATCTTCACTCACTGGATTCGGGTG 1345

Search completed: August 14, 2003, 11:38:24
Job time : 482.467 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:17:45 / Search time 3643.34 Seconds
(without alignments)
12848.224 Million cell updates/sec

Title: US-09-645-078-3
Perfect score: 1926
Sequence: 1 ggcgcactactactactg.....ccgtcacatctttctaag 1926

Scoring table: IDENTITY_NDC
Gapop 10.0, Gapext 1.0

Searched: 22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters: 45562784

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hlc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hlc:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrt:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	1499.8	77.9	1923	AK009113 Mus muscu
2	804.8	41.8	902	BQ947021 AGENCOURT
3	635.6	33.0	783	BG963298 602827716
4	628.4	32.6	965	BF579746 602095056

5	604.4	31.4	695	12	BG964671	BG964671 602831875
6	602.2	31.3	852	12	BG966340	BG966340 602832826
7	559	29.0	571	12	BM245312	BM245312 K0722907-
8	548.8	28.5	656	9	AA522184	AA522184 v19809.r
9	495	25.7	507	12	BM246681	BM246681 K0741E04-
10	434.2	22.5	954	12	B1823850	B1823850 603039012
11	353.2	18.3	668	14	BY708952	BY708952 BY708952
12	346	18.0	603	9	AL709927	AL709927 DKFZP6860
13	305	15.8	583	2	BSM082755	BSM082755 Homo sapi
14	292.4	15.2	571	12	BM129080	BM129080 H117C04.Y
15	267	13.9	834	13	BQ884877	BQ884877 AGENCOURT
16	249	12.9	604	13	B0135348	B0135348 603125426
17	247	12.8	895	10	BE857538	BE857538 7601A08.x
18	245.8	12.8	295	9	A1614882	A1614882 v19809.Y
19	238.2	12.4	692	14	BY711808	BY711808 BY711808
20	234.4	12.2	735	10	BE858652	BE858652 7601A09.x
21	228.4	11.9	620	9	A1824100	A1824100 w14601.x
22	225.8	11.7	904	13	B0241318	B0241318 603323578
23	221.4	11.5	773	13	B0458165	B0458165 603774342
24	217.4	11.3	703	14	CD218830	CD218830 pgrln.pk0
25	213.6	11.1	1067	29	CNS03KN7	AL248380 Tetradon
26	210.6	10.9	768	13	B0342120	B0342120 603517960
27	202.2	10.3	692	13	B0256241	B0256241 603746827
28	195.4	10.1	525	10	BF197521	BF197521 7684A08.x
29	187.2	9.7	516	9	AA689056	AA689056 v806C06.r
30	187	9.7	527	4	BX519156	BX519156 RZPD Mus
31	167	8.7	849	29	CNS04OEN	AL302540 Tetradon
32	165.8	8.6	568	9	A1115260	A1115260 u143c07.Y
33	159.4	8.3	543	13	B0382276	B0382276 603855935
34	153.8	8.0	555	4	BX514916	BX514916 RZPD Mus
35	151	7.8	517	10	BF042384	BF042384 BP250022A
36	148.6	7.7	657	29	AG035205	AG035205 Pan trogl
37	147.6	7.7	433	28	AZ405100	AZ405100 1M0173108
38	135.8	7.1	429	10	BB849113	BB849113 BB849113
39	131.8	6.8	758	13	B0259678	B0259678 603504181
40	131.6	6.8	599	12	B031352	B031352 B031352
41	129.8	6.7	436	9	AM081348	AM081348 xc41b06.x
42	126.2	6.6	498	13	BX280599	BX280599 BX280599
43	124	6.4	830	14	CB558413	CB558413 AGENCOURT
44	122.6	6.4	961	13	BQ734733	BQ734733 AGENCOURT
45	121.2	6.3	423	13	BY237139	BY237139 BY237139

ALIGNMENTS

RESULT 1
AK009113
LOCUS
DEFINITION Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310003618 product:carbohydrate (chondroitin 6/keratan) sulfotransferase 4, full insert sequence.
ACCESSION AK009113
VERSION AK009113.1 GI:12843701
KEYWORDS HTG; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
REFERENCE
1 Carninci, P., and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
MEDLINE 99279253
PUBMED 10349636
AUTHORS
TITLE
JOURNAL
MEDLINE
PUBMED
10349636
Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Kono, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
JOURNAL 20499374
MEDLINE 11042159
PUBMED 11042159


```

|||||
Db 572 CTTTGAATATGTTGAGAGAGCCCTGCGCTTCACAGGGTCTGTGTACTCAGAGGAGTGGC 631
QY 941 TTTTCTCAGGCTGAGGAGCCCTATCTACACTACAGGAGACCTTCCCTCAACGCGACGT 1000
Db 632 TTTTCTCAGGCTGAGGAGCCCTATCTACACTACAGGAGACCTTCCCTCAACGCGACGT 691
QY 1001 CGTGACACTGTTGAGAGAGCCCGGGGCGTGTCCGATGCCGAGAGACACCACTATAGA 1060
Db 692 CGTGACACTGTTGAGAGAGCCCGGGGCGTGTCCGATGCCGAGAGACACCACTATAGA 751
QY 1061 ACTCATGTTGAGAGATATTTGTGTAGGAGACACTTTGGAAACGATCAAGAGAGAGA 1120
Db 752 ACTCATGTTGAGAGATATTTGTGTAGGAGACACTTTGGAAACGATCAAGAGAGAGA 811
QY 1121 CGAGCCCTATTATGCGATGGAAGATCATCTGCAAAAAGCCAGGTGACATAGCAGGCCAT 1180
Db 812 CGAGCCCTATTATGCGATGGAAGATCATCTGCAAAAAGCCAGGTGACATAGCAGGCCAT 871
QY 1181 CCAAAACCTCCCTGAAAGCTGTGACAGAGCCCTACCTTCTCTGAGTATGAGACCTGGT 1240
Db 872 CCAAAACCTCCCTGAAAGCTGTGACAGAGCCCTACCTTCTCTGAGTATGAGACCTGGT 931
QY 1241 TCGGGGACCCCTGGCCGAGACGACAGACTATATAATTTGTGGGTTGATTTTGGC 1300
Db 932 TCGGGGACCCCTGGCCGAGACGACAGACTATATAATTTGTGGGTTGATTTTGGC 991
QY 1301 CGACCTCCAAACAGTGGTTTACAATGTCACCCCGGGGGAAGGCGATGGGTACAGATGCTT 1360
Db 992 CGACCTCCAAACAGTGGTTTACAATGTCACCCCGGGGGAAGGCGATGGGTACAGATGCTT 1051
QY 1361 CCATTAACAGCAGAGAGAGCCCTCAACGCTCTCTCAGAGCGTGGCGTGGTCTTACCTTA 1420
Db 1052 CCATTAACAGCAGAGAGAGCCCTCAACGCTCTCTCAGAGCGTGGCGTGGTCTTACCTTA 1111
QY 1421 CGAAAAGTTTCCAGCTTCAAGATGCTGCGGTGAGGCTATGATTTGCTGGGATACCT 1480
Db 1112 CGAAAAGTTTCCAGCTTCAAGATGCTGCGGTGAGGCTATGATTTGCTGGGATACCT 1171
QY 1481 CGAGGTCAGATTCACAAACAGAGCAAGCAACGCTGCTCCGATGCTTCTGCTCTCCCA 1540
Db 1172 CGAGGTCAGATTCACAAACAGAGCAAGCAACGCTGCTCCGATGCTTCTGCTCTCCCA 1231
QY 1541 TATCTTGGGAGAGTCTTCCGAGAGGTTAAGGAGCTGTGCTGACACCCCTTGGTTCAG 1600
Db 1232 TATCTTGGGAGAGTCTTCCGAGAGGTTAAGGAGCTGTGCTGACACCCCTTGGTTCAG 1291
QY 1601 CCTTAGTACCAATTTAAACGACAGAGCCCTTAAGTATAACCAAACTGAGTGGCCCTTTC 1660
Db 1292 CCTTAGTACCAATTTAAACGACAGAGCCCTTAAGTATAACCAAACTGAGTGGCCCTTTC 1351
QY 1661 TCCGTCAGCCCAAGCAGAGAGGAGTCTTGTCTATATCTCATGTCATCTCCCTCAACATGAGC 1720
Db 1352 TCCGTCAGCCCAAGCAGAGAGGAGTCTTGTCTATATCTCATGTCATCTCCCTCAACATGAGC 1411
QY 1721 CTAATAAGCCCAAGAAACAGTATCTTCTGTCTGTAATAATCTAGAACCTTAAGCAGCC 1780
Db 1412 CTAATAAGCCCAAGAAACAGTATCTTCTGTCTGTAATAATCTAGAACCTTAAGCAGCC 1471
QY 1781 CCTTAGACCTGTCAAGAAACATTTCTTGTAACTTGGCTTCTTACCTGTGATACCTT 1840
Db 1472 CCTTAGACCTGTCAAGAAACATTTCTTGTAACTTGGCTTCTTACCTGTGATACCTT 1511
QY 1841 GAGAGACT-CGGTGTGAGAGCATACTGAGACAGAGAAACAGATCTGGAGTGTGTGT 1899
Db 1532 GAGAGACTGAGGCTGTGAGAGCATACTGAGACAGAGAAACAGATCTGTGAGTGTGTGT 1591
QY 1900 AAACTCCTCCTGTACATCTTTTC 1922
Db 1592 AAACTCCTCCTGTACATCTTTTC 1614

```

```

B0947021
LOCUS B0947021 902 bp mRNA linear EST 21-AUG-2002
DEFINITION AGENCOURT 8951359 NCI CGAP C024 Mus musculus cDNA clone
IMAGE:6474928 5', mRNA sequence.
ACCESSION B0947021
VERSION B0947021.1 GI:22362499
KEYWORDS EST.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 902)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: The Cepko Laboratory
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LMNL)
DNA Sequencing by: Agencourt Bioscience Corporation
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LMNL at:
http://image.llnl.gov
Plate: LMNL4012 row: n column: 17
High quality sequence stop: 633.
FEATURES
location/qualifiers
SOURCE 1..902
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:6474928"
/lab_host="DH10B (TI phage-resistant)"
/clone_id="NCI_CGAP_C024"
/note="Organ: colon; Vector: pCMV-Sport6; site_1: NotI;
site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI CGAP Library."
BASE COUNT 228 a 248 c 210 g 213 t 3 others
ORIGIN
Query Match 41.8%; Score 804.8; DB 13; Length 902;
Best Local Similarity 97.4%; Pred. No. 1.5e-136;
Matches 827; Conservative 0; Mismatches 20; Indels 2; Gaps 1;
QY 1048 ACACCACCATGAAACATGATGTTGACAGTCAATTTGCTAGGAGCAGCATTTGAAACGA 1107
Db 1 ACACCACCATGAAACATGATGTTGACAGTCAATTTGCTAGGAGCAGCATTTGAAACGA 60
QY 1108 TCAGAGAGAGAGACAGCCCTATATGCGATGGAAGATCATGTGCAAAAAGCCAGGTGGACA 1167
Db 61 TCAGAGAGAGAGACAGCCCTATATGCGATGGAAGATCATGTGCAAAAAGCCAGGTGGACA 120
QY 1168 TAGTCAGAGCCATCAAAACCTCCCTGAAGCTCTGCAGCAGCGCTACCTGTTCTGAGGT 1227
Db 121 TAGTCAGAGCCATCAAAACCTCCCTGAAGCTCTGCAGCAGCGCTACCTGTTCTGAGGT 180
QY 1228 ATGAGGACCTGTGTTGGGAGACCCCTGGCCAGAGACAGACTATATTAATTTGTGGGCT 1287
Db 181 ATGAGGACCTGTGTTGGGAGACCCCTGGCCAGAGACAGACTATATTAATTTGTGGGCT 240
QY 1288 TGAATTTTGGCCCACTCCCAACATGAGGTTTCAATGTATACCCGGGCAAGGAGCATGG 1347
Db 241 TGAATTTTGGCCCACTCCCAACATGAGGTTTCAATGTATACCCGGGCAAGGAGCATGG 300
QY 1348 GTGAGATGCTTTCATATCAAGCCAGAGAAAGCCCTCAAGCTCTCAGGCGTGGCGTT 1407
Db 301 GTGAGATGCTTTCATATCAAGCCAGAGAAAGCCCTCAAGCTCTCAGGCGTGGCGTT 360
QY 1408 GGTCTTACCTTACGAAAAGTTTCCAGAGCTTCAAGATGCTGCGGTGAGGCTATGATTT 1467
Db 361 GGTCTTACCTTACGAAAAGTTTCCAGAGCTTCAAGATGCTGCGGTGAGGCTATGATTT 420

```

QY 1468 TGTCTGGATACCTCCAGCTCAGATCTCAACAGAACAGCAACCTGTCCCTGGATCTTC 1527
 DB 421 TGTCTGGATACCTCCAGCTCAGATCTCAACAGAACAGCAACCTGTCCCTGGATCTTC 480
 QY 1528 TGTCTCTCCCTCCATATCTTGGGGGAGGCTTCTCCGAGAGGTTAAGAGGCTGTCTGCAC 1587
 DB 481 TGTCTCTCTCCCTCCATATCTTGGGGGAGGCTTCTCCGAGAGGTTAAGAGGCTGTCTGCAC 540
 QY 1588 CCCTTGTTCAGACCTTGTAGTCAACATTAACGACAGAGGCTTGTAGTATACCAACT 1647
 DB 541 CCCTTGTTCAGACCTTGTAGTCAACATTAACGACAGAGGCTTGTAGTATACCAACT 600
 QY 1648 GAGTCCCTCTTCTCTCTAGCCGAGGAGGCTTGTGTCTATCTATCTCTCTAC 1707
 DB 601 GAGTCCCTCTTCTCTCTAGCCGAGGAGGCTTGTGTCTATCTATCTCTCTAC 660
 QY 1708 CCTACACTGAGCCTAAAGAGCAAGAAACAGTATCTTCTGTCTTGAATACTTAGA 1767
 DB 661 CCTACACTGAGCCTAAAGAGCAAGAAACAGTATCTTCTGTCTTGAATACTTAGA 720
 QY 1768 ACCTTAAGAGCCCTTGTACCTGTCAAGCAAGACTTCTGTACCTTGGCTTCTAC 1827
 DB 721 ACCTTAAGAGCCCTTGTACCTGTCAAGCAAGACTTCTGTACCTTGGCTTCTAC 780
 QY 1828 CTGTGCATACCTTGTAGACT--CGTCTGAGGACATCTGACACAGCAAGACTCTG 1885
 DB 781 CTGTGCATACCTTGTAGACTTGTGTGAGGAGGCTGTGAGGAGGCTGTGAGCAACANCA 840
 QY 1886 TGGAGTGTG 1894
 DB 841 TGGAGAGTG 849
 RESULT 3
 LOCUS BG963298 783 bp mRNA linear EST 12-JUN-2001
 DEFINITION 602827716F1 NCI_GAP_Co24 Mus musculus cDNA clone IMAGE:4982643 5',
 mRNA sequence.
 ACCESSION BG963298.1 GI:14350935
 VERSION BG963298.1
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 783)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgapbs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 plate: ILAM10985 row: P column: 04
 High quality sequence stop: 741.
 Location/Qualifiers
 1..783
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone_image="4982643"
 /lab_host="DH10B (TI phage-resistant)"
 /clone_lib="NCI_GAP_Co24"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site: 1; NotI;
 Site: 2; SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.6 kb. Constructed by Life

BASE COUNT 141 a 236 c 226 g 180 t
 ORIGIN
 Query Match 33.0%; Score 635.6; DB 12: Length 783;
 Best Local Similarity 98.0%; Pred. No. 9.6e-106;
 Matches 656; Conservative 0; Mismatches 9; Indels 5; Gaps 5;
 QY 398 CCGCAGATGATGCTGTGAAGAAAGAGGCTGCTGATGTTCTCGGTTCCAGTCA 457
 DB 60 CCTAAGATGATGCTGTGAAGAAAGAGGCTGCTGATGTTCTCGGTTCCAGTCA 119
 QY 458 CGTTGAGCTCTCTTCATCATATATGCTCCACAGACACTTCCACAGAGAGAGAC 517
 DB 120 CGTTGAGCTCTCTTCATCATATATGCTCCACAGACACTTCCACAGAGAGAGAC 179
 QY 518 CAGAGAGCCGTCATGTCGTGTCTCTCTCTGAGAGATCTCTTTGTGG 577
 DB 180 CAGAGAGCCGTCATGTCGTGTCTCTCTCTGAGAGATCTCTTTGTGG 239
 QY 578 ACAGCTTTTCGGGACACACCCGAGATGTTCTACCTGATGAGACCTGCTGGCAGTGTG 637
 DB 240 ACAGCTTTTCGGGACACACCCGAGATGTTCTACCTGATGAGACCTGCTGGCAGTGTG 299
 QY 638 GATGACTTTCACACAGACACACCCGAGATGTTCTACCTGATGAGACCTGCTGGCAGTGTG 697
 DB 300 GATGACTTTCACACAGACACACCCGAGATGTTCTACCTGATGAGACCTGCTGGCAGTGTG 359
 QY 698 TTCCGCTCTCTGTGATGATGAGCTCTTTGATGCTTACATGAAACCCAGGCCCGGAA 757
 DB 360 TTCCGCTCTCTGTGATGATGAGCTCTTTGATGCTTACATGAAACCCAGGCCCGGAA 419
 QY 758 ACAGTCCAGCTCTTCACTGAGGAGCAAGCCGAGCTGTGCTACAGGCTGTGTGTA 817
 DB 420 ACAGTCCAGCTCTTCACTGAGGAGCAAGCCGAGCTGTGCTACAGGCTGTGTGTA 479
 QY 818 CTTCCTCCCTGCGCAAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 877
 DB 480 CTTCCTCCCTGCGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 539
 QY 878 GCCCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 936
 DB 540 GCCCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 599
 QY 937 TGGCT-TTTCCTAGCTGAGGAGGCTGTATTCACACTACTAGGAGACCTTCCCTCAACTG 995
 DB 600 TGGCT-TTTCCTAGCTGAGGAGGCTGTATTCACACTACTAGGAGACCTTCCCTCAACTG 659
 QY 996 -CACGTCGACACCTGTCGAGAGACCCGAGGCTGTCGATCCCGGAGACACAC 1054
 DB 660 CACGTCGACACCTGTCGAGAGACCCGAGGCTGTCGATCCCGGAGACACAC 718
 QY 1055 CATAGACACTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1103
 DB 719 CATAGACACTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 768
 RESULT 4
 LOCUS BF579746 965 bp mRNA linear EST 12-DEC-2000
 DEFINITION 602095056F1 NCI_GAP_Co24 Mus musculus cDNA clone IMAGE:4209636 5',
 mRNA sequence.
 ACCESSION BF579746
 VERSION BF579746.1 GI:11653458
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 965)
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished

QY 620 GCTTGGCTGGATGTGTGATGACTTTTCACGACGACAGCCTGGAGCTGACATGAC 679
 DB 251 GCTTGGCTGGATGTGTGATGACTTTTCACGACGACAGCCTGGAGCTGACATGAC 310
 QY 680 TGTGGGGGATCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 739
 DB 311 TGTGGGGGATCTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 370
 QY 740 GAACCCAGGCCCCGGAAACAGTCCAGCCTTTCAGTGGAGCAAGCCGGGCTGTG 799
 DB 371 GAACCCAGGCCCCGGAAACAGTCCAGCCTTTCAGTGGAGCAAGCCGGGCTGTG 430
 QY 800 CTCAGGCGCTGTGTGACTTCTTCCCTGCGCAGAGATCAGCTACCCAGACATGCAA 859
 DB 431 CTCAGGCGCTGTGTGACTTCTTCCCTGCGCAGAGATCAGCTACCCAGACATGCAA 490
 QY 860 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 919
 DB 491 GGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 550
 QY 920 CGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 979
 DB 551 CGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 610
 QY 980 CCGCTT-CCCTCAACCTGACAGT-CGTGACAGCTGCTGCTGCTGCTGCTGCTGCTGCT 1037
 DB 611 CCGCTTCCCTCAACCTGACAGTCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 669
 QY 1038 TCCCGGGAGACAC 1051
 DB 670 TCCCGGGAGACAC 683

RESULT 6
 BG966340 852 bp mRNA 1linear EST 12-JUN-2001
 LOCUS 602832826F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4987327 5'
 DEFINITION mRNA sequence.
 ACCESSION BG966340
 VERSION BG966340.1 GI:14353977
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgc.ncl.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaabp-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Inqye Genomics, Inc.
 Clone distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L14M10998 row: c column: 08
 High quality sequence stop: 784.
 Location/Qualifiers
 1..852
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4987327"
 /lab_host="DH10B (TI phage-resistant)"
 /clone_lib="NCI_CGAP_Co24"
 /note="Organ: colon; Vector: pcMV-SPO6; Site:1; NotI;
 Site:2; SalI; Cloned unidirectionally. Primer: Oligo dT.
 Average insert size 1.6 kb. Constructed by Life

FEATURES
 source

BASE COUNT 176 a 248 c 244 g 184 t
 ORIGIN
 Query Match 31.3%; Score 602.2; DB 12: Length 852;
 Best Local Similarity 96.4%; Pred. No. 1.1e-99;
 Matches 670; Conservative 0; Mismatches 18; Indels 7; Gaps 5;

QY 401 CAGATGATGCTGCTTGAAGAAAGAGGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 460
 DB 133 CAGATGATGCTGCTTGAAGAAAGAGGCTGCTGATGCTGCTGCTGCTGCTGCTGCTGCT 192
 QY 461 TGTAGCTCTCTTCATCATATGTCGCTCCACAGACACCTTTCACAGAGGAGATCCAG 520
 DB 193 TGTAGCTCTCTTCATCATATGTCGCTCCACAGACACCTTTCACAGAGGAGATCCAG 252
 QY 521 GAGGCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 579
 DB 253 GAGGCGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 312
 QY 580 AGCTTTCCGCGCAGCACCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 639
 DB 313 AGCTTTCCGCGCAGCACCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 372
 QY 640 TGACTTTCAACAGCAGACAGCCTGAGACCTGACATGCTGCTGCTGCTGCTGCTGCTGCT 699
 DB 373 TGACTTTCAACAGCAGACAGCCTGAGACCTGACATGCTGCTGCTGCTGCTGCTGCTGCT 432
 QY 700 CCGTCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 759
 DB 433 CCGTCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 492
 QY 760 AGTCCAGCCTCTTCCAGTGGAGCAAGCCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 819
 DB 493 AGTCCAGCCTCTTCCAGTGGAGCAAGCCGGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 552
 QY 820 TCTTCCCTGCCACGAGATCAGCTACCCAGACATGCAAGCTGCTGCTGCTGCTGCTGCT 879
 DB 553 TCTTCCCTGCCACGAGATCAGCTACCCAGACATGCAAGCTGCTGCTGCTGCTGCTGCT 612
 QY 880 CCTTGTATATGTTGAGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 939
 DB 613 CCTTGTATATGTTGAGGAGAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 672
 QY 940 GTTTCCTGACCTGACAGGCTCTTATCCACTACTACAGACCTT-TCCTCAACCTG--C 996
 DB 673 GTTTCCTGACCTGACAGGCTCTTATCCACTACTACAGACCTTATCCACTCAACCTGGA 732
 QY 997 AGCTGCTGCA-CCTGCTCCGAGACCCCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1055
 DB 733 AGCTGCTGCA-CCTGCTCCGAGACCCCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 792
 QY 1056 ATGAACTCATGTTGACAGTATATGTTGCTAGG 1090
 DB 793 AAGAAATGCA--GGTGACCAATATATGTTGCTAGG 825

RESULT 7
 BM245312 571 bp mRNA 1linear EST 07-JUN-2003
 LOCUS BM245312
 DEFINITION K0722H07-3 NIA Mouse Hematopoietic Stem Cell (Lin-/c-Kit-/Sca-1-)
 CDNA Library (Long) Mus musculus cDNA clone NIA:K0722H07
 IMAGE:30075738 3', mRNA sequence.
 ACCESSION BM245312
 VERSION BM245312.2 GI:31500552
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 571)
 Reference
 Piao, Y., Ko, N.T., Lim, M.K. and Ko, M.S.H.
 Construction of long-transcript enriched cDNA libraries from

submicrogram amounts of total RNAs by a universal PCR amplification method
 Genome Res. 11 (9), 1553-1558 (2001)
 MEDLINE 21429098
 PUBMED 11544199
 On Dec 17, 2001 this sequence version replaced gi:17880582.
 Other ESTs: K0722H07-5N
 Contact: Dawood B. Dudekula
 Laboratory of Genetics
 National Institute on Aging/National Institutes of Health
 333 Cassell Drive, Suite 4000, Baltimore, MD 21224-6820, USA
 Email: cdna@igsun.igrs.nia.nih.gov
 Plate: K0722 Row: H Column: 07
 Seq primer: -21M13 Forward
 High quality sequence stop: 571
 POLY-A-Yes.

FEATURES

SOURCE

Location/Qualifiers
 1..571
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6NCr"
 /db_xref="tae:K0722H07-3"
 /db_xref="taxon:10090"
 /clone="NIA:K0722H07 IMAGE:30075738"
 /tissue_type="Hematopoietic Stem Cell (Lin-/C-Kit-/Sca-1-)"
 /dev_stage="Age approx.10 weeks old"
 /lab_host="DHI08"

(Lin-/C-Kit-/Sca-1-) cDNA library (Long)
 /note="Vector: pSPORT1 (Invitrogen); Site_1: SalI; Site_2: NotI; Mouse cDNA project by the Laboratory of Genetics, National Institute on Aging (NIA), Intramural Research Program, NIH (http://igsun.igrs.nia.nih.gov/cDNA). This is a long-transcript enriched cDNA library (Ref. Genome Res. 11: 1553-1558 (2001). PMID: 11544199). Total RNAs were obtained from Drs. Dennis Taub, Dan Longo (National Institute on Aging, USA), Jonathan Kessler (National Cancer Institute, USA). Double-stranded cDNAs were synthesized with an oligo(dT) primer (Invitrogen): 5'-pgactgttctgagatccgacgagccgccccttttttttttt-3' from 0.9 ug of total RNA, treated with T4 DNA polymerase, and purified by ethanol-precipitation. The cDNAs were ligated to Lona-linker L1-SalI, purified by phenol/chloroform, and separated from free linkers by Centricon 100. Then, the cDNAs were amplified by long-range high fidelity PCR using Ex Taq polymerase (Takara) with a primer Sal4-S. The products were purified by phenol/chloroform and Centricon 100. The cDNAs were digested with SalI and NotI enzymes and cloned into SalI/NotI site of pSPORT1 plasmid vector. The DHI08 E. coli host was transformed with the ligation mixture by the standard chemical method. The average insert size is about 2.1 kb. The library was constructed by Yulan Plao (NIA)."

BASE COUNT 98 a 162 c 155 g 156 t
 ORIGIN

Query Match 29.0%; Score 559; DB 12; Length 571;
 Best Local Similarity 99.8%; Pred. No. 8.8e-92;
 Matches 570; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

QY 275 GTTGTGTTGTTGTTTACATGCTGCTGGATGGAATCCA-GAGAAGCCCGAAGTAG 333
 Db 1 GTTGTGTTGTTGTTTACATGCTGCTGGATGGAATCCAAGAGTAG 60
 QY 334 ATGCTGAACACCTAATCAGCCCATCCCTGCTGCTCTTCAAGGCTCTTCCT 393
 Db 61 ATGCTGAACACCTAATCAGCCCATCCCTGCTGCTCTTCAAGGCTCTTCCT 120
 QY 394 TCTTCCGAGGATGCTGTTGAAGAAGGAGGCTGATGTTCTGCTGCTCCAGG 453
 Db 121 TCTTCCGAGGATGCTGTTGAAGAAGGAGGCTGATGTTCTGCTGCTCCAGG 180

QY 454 TCATCGTTGAGCTCTCTTCATTCATATGTCCTCACAGACACTTTCAGAGGAGG 513
 Db 181 TCATCGTTGAGCTCTCTTCATTCATATGTCCTCACAGACACTTTCAGAGGAGG 240
 QY 514 AGTCCAGAGAGCCGCTCATGCTGCTGCTGCTCTTCCGCTCAGATCTCTTTTG 573
 Db 241 AGTCCAGAGAGCCGCTCATGCTGCTGCTGCTCTTCCGCTCAGATCTCTTTTG 300
 QY 574 TGGACAGCTTTTGGGACAGACCCGATGCTTCACATGATGATGACCCGCTGGATG 633
 Db 301 TGGACAGCTTTTGGGACAGACCCGATGCTTCACATGATGATGACCCGCTGGATG 360
 QY 634 TGTGATGATCTTTCACAGAGACAGACCTGGAATGATGATGATGATGATGATGATG 693
 Db 361 TGTGATGATCTTTCACAGAGACAGACCTGGAATGATGATGATGATGATGATGATG 420
 QY 694 TGCCTCCGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 753
 Db 421 TGCCTCCGCTCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 480
 QY 754 GGAACAGTCCAGCTCTTCCAGTGGGAGCAAGCCGCGCTGCTGCTGCTGCTGCTGCT 813
 Db 481 GGAACAGTCCAGCTCTTCCAGTGGGAGCAAGCCGCGCTGCTGCTGCTGCTGCTGCT 540
 QY 814 GTGACTTCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 844
 Db 541 GTGACTTCTTCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 571

RESULT 8
 AA522184
 LOCUS
 DEFINITION
 v998b09.t1 Soares_mammary_gland_NbMNG Mus musculus cDNA clone
 IMAGE:851801 5', mRNA sequence.
 ACCESSION
 AA522184
 VERSION
 AA522184.1 GI:2262929
 KEYWORDS
 EST.
 SOURCE
 Mus musculus (house mouse)
 ORGANISM
 Mus musculus

REFERENCE
 AUTHORS
 Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T., Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M., Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B., Theising,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and Waterston,R.
 TITLE
 The WashU-HMI Mouse EST Project
 JOURNAL
 Unpublished
 COMMENT
 Contact: Marra M/Mouse EST Project
 WashU-HMI Mouse EST Project
 Washington University School of Medicine
 4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108
 Tel: 314 286 1800
 Fax: 314 286 1800
 Email: mouseest@wustl.edu
 This clone is available royalty-free through LNL; contact the IMAGE Consortium (info@image.llnl.gov) for further information.
 MGI:503953
 Seq primer: -28m13 rev2 ET from Amersham
 High quality sequence stop: 360.

FEATURES

SOURCE

Location/Qualifiers
 1..656
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="C57BL/6J"
 /db_xref="taxon:10090"
 /clone="IMAGE:851801"
 /sex="male"
 /tissue_type="mammary gland"
 /dev_stage="4 weeks"
 /lab_host="DHI08"
 /clone_lib="Soares_mammary_gland_NbMNG"

Db	121	ACTTCGCCGAGGATGATGCTGTTTGAAGAAAGGAGAGGCTGCTGATATGTTCTCGGGTTCCACAG	180
QY	454	TCATCGTTGTAGCTCTCTTCATCCATATGTCGTCACAGACACCTTTCCACAGAGGAGG	513
Db	181	TCATCGTTGTAGCTCTCTTCATCCATATGTCGTCACAGACACCTTTCCACAGAGGAGG	240
QY	514	AGTCAGGAGGAGCCGTGCATGATGCGTGGTGCCTTCTGCGGGTCAGAGATCCTCTTTTG	573
Db	241	AGTCAGGAGGAGCCGTGCATGATGCGTGGTGCCTTCTGCGGGTCAGAGATCCTCTTTTG	300
QY	574	TGGGACACCTTTTGGGCGCAGCACCCGGATGTTTCTACCTGATGAGAGCCTGCCTGGCATG	633
Db	301	TGGGACACCTTTTGGGCGCAGCACCCGGATGTTTCTACCTGATGAGAGCCTGCCTGGCATG	360
QY	634	TGTGGATGACTTTTACCACAGACACAGCCTGGAAAGCTGCATGAGCTGTGCGGAGATCTTTC	693
Db	361	TGTGGATGACTTTTACCACAGACACAGCCTGGAAAGCTGCATGAGCTGTGCGGAGATCTTTC	420
QY	694	TGCGTTCCGTTCTTCTGTGTGACATGAGCGTTTGTGATGCTCATATGAACCCAGGCCCC	753
Db	421	TGCGTTCCGTTCTTCTGTGTGACATGAGCGTTTGTGATGCTCATATGAACCCAGGCCCC	480
QY	754	GGAACAGATCCAGCCTTTCCAGTGGG	780
Db	481	GGAACAGATCCAGCCTTTCCAGTGGG	507

	RESULT 10
LOCUS B1823850	
DEFINITION B1823850 954 bp mRNA linear EST 04-OCT-2001 603039012P1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179826 5' , mRNA sequence.	
ACCESSION B1823850	
VERSION B1823850.1 GI:15935400	
KEYWORDS EST..	
SOURCE Homo sapiens (human)	
TAXIDOM Homo sapiens	
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Euarcharia; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 954) NIH-MGC http://mgc.nci.nih.gov/. National Institutes of Health, Mammalian Gene Collection (MGC) Unpublished	
REFERENCE Contact: Robert Strausberg, Ph.D. Email: rsapbs@mail.nih.gov Tissue Procurement: Life Technologies, Inc. CDNA Library Preparation: Life Technologies, Inc. CDNA library Arrayed by: The I.M.A.G.E. Consortium (LMNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LMNL at: http://image.lnl.gov	
AUTHORS Plate: LMAM1448 row: h column: 03	
JOURNAL High quality sequence stop: 856.	
COMMENT	

FEATURES

source

```

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5179826"
/lab_host="DH10B"
/clone_1fb="NH_MGC_115"
/note="Organ: pooled brain, lung, testis; Vector:
pCMV-SPORT6; Site:1: NotI; site_2: EcoRV (destroyed); RNA
source anonymous pool of 6 male brains, age 23-27; 1
male lung, age 27; and 1 male testis, age 65. Library is
oligo-dT primed and directionally cloned (EcoRV site is
destroyed upon cloning). Average insert size 1.8 kb,
insert size range 1-3 kb. Library is normalized and
enriched for full-length clones and was constructed by C.
Gruber (Invitrogen). Research Genetics tracking code
021. Note: this is a NH_MGC library."
BASE COUNT
199 a 297 c 250 g 208 t

```

Query Match	22.5%	Score 434.2;	DB 12;	Length 954;
Best Local Similarity	73.9%;	Pred. No. 4.2e-69;		
Matches 660; Conservative	0;	Mismatches 218;	Indels 15;	Gaps 8

OY	388	CTCTCTCTCCGAGAGATGATGCGTGTGAAGAAAGGAGAGCGTGTATGTTCCCGGGT	447
Db	61	TCCTCCACTCAGCACAATGCTACTGCTTAAAAAATTAATAGCTCCTGCTGTTCTGGTTT	120
OY	448	CCAGAGTCATCGTTGTAGCTCTCTTCATCCATATGTCCGCACAGACACTTTCAG- 	506
Db	121	CCGAGATGGCCATCTTGSGCTCTATTCTTCCACATGTATACAGCCACAACATCAGCTCCGT	180
OY	507	--AGGAGAGTCCAGAGGCCCGGTGCATGTGCTGTGCTCTTCTTGCGCGTATAGAT 	564
Db	181	CTATGAAAGGACACAGGCCGAGCGCATGACAGTGCCTGTTCTGTCCTCGAGCGCTCGGCT	240
OY	565	CCTCTTGTGTGGGACAGCTTTTTCGGGCAAGCACCAGATGTGTTCTTACTGATGAGCCTG	624
Db	241	CTTCTTTTGTGGGCGAGCTTTTTCGGGCAAGCACCAGATGTGTTCTTACTGATGAGCCTG	300
OY	625	CCTGGCATGTGAGATGACTTTCACACAGACAGCCTGGAAAGTGTGACATGAGTGTGC 	684
Db	301	CCTGGCAGCTGTGAGATGACTTTCACAGCAGAGACACCGCCTGGATGTGACACATGSGTGTGC	360
OY	685	GGAGATCTTCTGCGTTCCGTCCTTCGATGTGACATAGAGTCTTTTATGCTCTTACATGAAC 	744
Db	361	GGGATCTATACGGGCGCTTCTTCTGTGTGCGACATAGCGCTTTTATGCTCTTACATGAAC	420

OY	745	CAGGCCCCGGAAACAGTCACAGCCCTCTTCCAGTGGAGCAAAAGCCGGGCCCTGTGCTCAG	804
Db	421	CTGGGCCCCGGAGAGACAGTCCAGCCCTCTTCCAGTGGAGAAACAGCCGGGCCCTGTGTCTG	480
OY	805	CGCCGTGTGTGACCTCTTCCCTGCGCCAGAGATAGCTACGCCCAAGACACGTGACAGCTGC	864
Db	481	CACCTGCTGTGACATCATCCACAGATGAATATATCCCCGGGCTCACTGACAGGCTCC	540
OY	865	TCTGCGGTACAGAGCCCTTGTGATATGATGGAGAAAGGCCCTGCGGCTCTACAGGCTTCTG	924
Db	541	TGTGACGTCAACAGGCCCTTGTGAGTGGTGGAGAAAGGCCCTGCGGCTCTTACAGCACGTGG	600
OY	925	TACTCAAGAGAGTGGCTTTTCTCAGCCTTCAGAGGCCCTCTATCTACTACTACAGGACCTT	984
Db	601	TGCTCAAGAGAGTGGCTTCTTCAACCTTCGACATCTCCCTCTTACCCGGTGTGAAAGACCCCT	660
OY	985	CCCTCAACCTCAGCTCGCTGCACAC - TG6TTCGAGACCCCGGGGCGGTTCGCATCCCGG	1045
Db	661	CCCTCAACCTGCATTCGTGTGCACCTTGGTCCGGGAGCCCGGGCGGTGTCCGTTCCCGGA	720
OY	1044	GAGCACACC - ACCATAGAACTCATGTTGA - CAGTCATATTGTGCTAGGGCAGCAATTGG	1101
Db	721	GAAGCGACCAAAAGGAGATCTCATGATTAGCACAGTCGATGTGTGGGGCAGCATGTAGC	780
OY	1102	AAA-----CGATCAAGAGAGGAAGACAGCCCTTATATGATGCATGAAGATCATCTGCAAAAG	1156
Db	781	AAAAAACTCAAGAGAGAGAGAGACCAACCTTACTATGTATGTGACAGGTATCTGCCAAAG	840
OY	1157	CCAGGTGACATAGT - CAAGGCCATCCAAACCCCTCCCTGAAGCTCTGCACAGACGGCTAC	1214
Db	841	CCAGGCTGTGAATCTATCAAAAGCAACCTTCAAGTCTTGGCCCAAGGGCCCTCAGAGAACGTAC	900
OY	1215	CTGTTTCCCTGAG - GTATGAGGAGCACTGTTTCG - GGACACCCCTGGGCCCAAGAGACC	1265
Db	901	CTGCTTGTGCGCAATGAGGACCTTGGTTCGAGGCCCCCTGTGGCCACGACTTCC	953

VERSION
 BY0708952.1 GI:27120147
 EST.

KEYWORDS
 Mus musculus (house mouse)

SOURCE
 Mus musculus

ORGANISM
 Mus musculus (house mouse)

REFERENCE
 Authors
 ENKAYIYOLA, Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus. 1 (bases 1 to 803)
 Okazaki, Y., Furuno, M., Kasukawa, T., Adachi, J., Bono, H., Kondo, S., Nishida, I., Otsu, N., Matsuura, K., Saito, R., Suzuki, H., Yamane, I., Kiyosawa, H., Yagi, K., Tomaru, Y., Hasegawa, Y., Nogami, A., Schombach, C., Gotohori, T., Baldarelli, R., Hill, D. P., Bult, C., Hume, D. A., Quackenbush, J., Schmitt, L. M., Knapik, A., Matsuda, H., Batilov, S., Beisel, K. W., Blake, J. A., Brad, D., Brusic, V., Chothia, C., Corbani, L. E., Conlins, S., Dalla, E., Dragani, T. A., Fletcher, C. F., Forrest, A., Fraser, R. S., Gaasterland, T., Gariboldi, M., Gissi, C., Godzik, A., Gough, P., Grimmond, S., Gustincich, S., Hirokawa, N., Jackson, I. J., Jarvis, E. D., Kanai, A., Kawaji, H., Kawasawa, Y., Kedziarski, R. M., King, B. L., Konagaya, A., Kurochkin, I. V., Lee, Y., Lenhard, B., Lyons, P. A., Maglott, D. R., Matis, L., Marchionni, L., McKenzie, L., Mikl, H., Nagashima, T., Numata, K., Okido, T., Pavan, W. J., Pettes, G., Pesole, G., Petrovsky, N., Pillai, R., Pontus, J. V., Qi, D., Ramachandran, S., Ravasi, T., Reed, J. C., Reed, D. J., Reid, J., Ring, B. Z., Ringwald, M., Sandelin, A., Schneider, C., Semple, C. A., Setou, M., Shimada, K., Sultana, R., Takenaka, Y., Taylor, M. S., Teasdale, R. D., Tomita, M., Verrardo, R., Wagner, L., Wahlestedt, C., Wang, Y., Watanabe, Y., Wells, C., Wilming, L. G., Wyman, B. J., Yang, A., Yang, A., Yang, I., Yang, L., Yuan, Z., Zavolan, M., Zhu, Y., Zimmer, A., Carninci, P., Hayatsu, N., Hirozane-Kitahara, T., Kono, H., Nakamura, M., Sakazume, N., Sato, K., Shiraki, T., Waki, K., Iwami, J., Aizawa, K., Arakawa, T., Fukuda, S., Hara, A., Hashizume, W., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Miyazaki, A., Sakai, K., Sasaki, D., Shibata, K., Shingawa, A., Yasunishi, A., Yoshino, M., Waterston, R., Lander, E. S., Rogers, J., Birney, E., and Hayashizaki, Y.
 Analysis of the mouse transcriptome based on functional annotation
 of 60,770 full-length cDNAs
 Nature 420, 563-573 (2002)

TITLE
 Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs

JOURNAL
 MEDLINE
 PUBMED
 22354683
 12466851

COMMENT
 Contact: Yoshihide Hayashizaki
 Laboratory for Genome Exploration Research Group, RIKEN Genomic Sciences Center (GSC), Yokohama Institute
 The Institute of Physical and Chemical Research (RIKEN)
 1-7-22 Suhiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
 Tel: 81-45-503-9222
 Fax: 81-45-503-9216
 Email: genome-resgsc.riken.go.jp
 URL: http://genome.gsc.riken.go.jp/
 Adachi, J., Aizawa, K., Akimura, T., Arakawa, T., Carninci, P., Fukuda, S., Hashizume, W., Hayashida, K., Hirozane, T., Hori, F., Imotani, K., Ishii, Y., Itoh, M., Kagawa, I., Kawaji, D., Kojima, Y., Kondo, S., Kono, H., Koya, S., Miyazaki, A., Murata, M., Nakamura, M., Nomura, K., Nunazaki, R., Ohno, M., Ohsato, N., Saito, R., Sakazume, N., Sano, H., Sasaki, D., Sato, K., Shibata, K., Shiraki, T., Tsegami, M., Takeda, Y., Waki, K., Watanabe, A., Muramatsu, M., and Hayashizaki, Y. Direct Submission
 Computational Analysis of Full-length Mouse cDNAs Compared with Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
 Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes. Genome Res. 10 (10), 1617-1630 (2000)
 RIKEN Integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multicapillary sequencer. Genome Res. 10 (11), 1757-1771 (2000)
 Computer-based methods for the mouse full-length cDNA encyclopedia: real-time sequence clustering for construction of a nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
 cDNA library was prepared and sequenced in Mouse Genome Encyclopedia Project of Genome Exploration Research Group in Riken Genomic Sciences Center and Genome Science Laboratory in Riken. Division of Experimental Animal Research in Riken contributed to prepare mouse tissues.
 Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

JOURNAL COMMENT

Unpublished
Contact: Bahr A
MIPS

Ingolstaedter Landstr.1, D-85764 Neuherberg, Germany
This is the 5' sequence of the clone insert
clone from S. Wiemann, Molecular Genome Analysis, German Cancer
Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
sequenced by Olagen (Hilden/Germany) within the cDNA sequencing
consortium of the German Genome Project.

No sl sequence available.

This clone (DKFZp68602364) is available at the RZPD in Berlin.
Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.

FEATURES

source

1.668
Location/Qualifiers

/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="DKFZp68602364"
/tissue_type="human skeletal muscle"
/dev_stage="adult"
/lab_host="DH10B"
/clone_id="686 (synonym: hlcc3)"
/note="Vector: pTRIPlex2; Site_1: sf11A; site_2: sf11B;
cDNA-collection"

BASE COUNT 132 a 202 c 178 g 154 t 2 others
ORIGIN

Query Match 18.0%; Score 346; DB 9; Length 668;
Best Local Similarity 76.2%; Pred. No. 4.9e-53;
Matches 451; Conservative 0; Mismatches 137; Indels 4; Gaps 2;

388 TCTCCCTCTCCGAGATGCTGTGTAAGAAAGGAGGCTGCTATGTCCTGGCT 447
Db TCTTCCACTTCAGACATCTACTGCTTAAATAAGAGCTCTGCTGTGCTT 136
448 CCCAGGTCATCTGTAGCTCTCTCATTCATATGTCGCCACAGACACTTCCAG- 506
Db CCCAGATGGCCATCTTGCTCTATCTTCCACATGTACACCAACATCAGCCCTGT 196
137 CCCAGATGGCCATCTTGCTCTATCTTCCACATGTACACCAACATCAGCCCTGT 196
507 --AGGAGAGAGTCCAGAGGCCCGCTGATGCTGCTGCTCTTCTGCGCTGAGAT 564
Db CTATGAGAGGACAGAGCCGAGCGATGACGCTGCTGCTCTGCTGCTGCTGCT 256
197 CTATGAGAGGACAGAGCCGAGCGATGACGCTGCTGCTCTGCTGCTGCTGCT 256
565 CTTCTTTTGGGAGACACTTTTGGGAGACACCGGATGTTCTTACATGAGAGCTG 624
Db CTTCTTTTGGGAGACACTTTTGGGAGACACCGGATGTTCTTACATGAGAGCTG 624
257 CTTCTTTTGGGAGACACTTTTGGGAGACACCGGATGTTCTTACATGAGAGCTG 316
625 CTTGGCATGTGTGATGATCTTTCACAGACAGACAGCTGGAAGCTGACATGGCTGTC 684
Db CTTGGCATGTGTGATGATCTTTCACAGACAGACAGCTGGAAGCTGACATGGCTGTC 684
317 CTTGGCATGTGTGATGATCTTTCACAGACAGACAGCTGGAAGCTGACATGGCTGTC 376
685 GGGATCTTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 744
Db GGGATCTTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 744
377 GGGATCTTCTGCTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 436
745 CAGGAG 804
Db CAGGAG 804
437 CAGGAG 496
805 CGCCTGTGTGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 863
Db CGCCTGTGTGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 863
497 CACCTGCTGTGATCATCCACAAAGATGAATCATCCCGGAGCTGCTGAGAGCTG 556
864 CTTGGGAGTCCAG 923
Db CTTGGGAGTCCAG 923
557 CTTGGGAGTCCAG 616
924 GTCCTAAG 975
Db GTCCTAAG 616
617 GTCCTAAG 668

RESULT 13
ID HSM082755 standard; RNA; EST; 583 BP.

AC BX490456;

SV BX490456.1

DT 09-MAY-2003 (Rel. 75, Created)
DT 09-MAY-2003 (Rel. 75, Last updated, Version 1)

DE Homo sapiens mRNA; EST DKFZp6860680_r1 (from clone DKFZp6860680)

XX EST; expressed sequence tag.

XX Homo sapiens (human)
XX Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
XX Eutheria; Primates; Catarrhini; Hominidae; Homo.

XX [1]

RP 1-583
RA Ansoer W., Krieger S., Regiert T., Rittmuller C., Schwager B.,
RA Mewes H.W., Well B., Amlid C., Osanger A., Fob G., Han M., Wiemann S.;

RL Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.
RL MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY

CC This is the 5' sequence of the clone insert
CC Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
CC Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
CC sequenced by EMBL (European Molecular Biology Laboratories),
CC Heidelberg/Germany) within the cDNA sequencing consortium of
CC the German Genome Project.

CC No sl sequence available.
CC This clone (DKFZp6860680) is available at the RZPD in Berlin.
CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,
CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

THIS PAGE BLANK (USPTO)


```

Db      | 345 | CTTCTTTTGTGGGAGCTTTTGGGAGCACCAGATGTTTCTACCTGATGGAGCCG 404
Oy      | 625 | CCTGCGATGTTGGATGATCTTTCACACAGCAGCAGCTGGAGAGCTGACATGGCTGTC 684
Db      | 405 | CCTGGCAGCTGTGATGATGCTTCAAGAGCAGCAGCAGCTGGATGCTGACATGGCTGTC 464
Oy      | 685 | GGGATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 744
Db      | 465 | GGGATCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 524
Oy      | 745 | CAGGCCCCGGGAACAGCAGCAGCTTCCAGTGGGAGCAAGCCGGGCGCTGCTGCTG 804
Db      | 525 | CTGGTCCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAGCAAGCCGGGCGCTGCTG 584
Oy      | 805 | CGCCTGTGTGTGATCTTCTTCCCTGCGCCAGAGATGCTCAGCCAGCAGTCAAGCTG 864
Db      | 585 | CACCTGCTGTGATCATCTCCACAGATGAAATCATCTCCCGGGGCTCAGTCAAGCTG 644
Oy      | 865 | TCTGCGGTGACAGCAGCCTTTGATATGCTGAGAAAGGCGCTGCTGCTGCTGCTG 924
Db      | 645 | TGTGAGTCAACAGCAGCCTTTGAGGTGTGAGAAAGGCGCTGCTGCTGCTGCTG 704
Oy      | 925 | TACTCAAGAGAGTGGTGTCTTCTGAGCCTGAGGCGCTTATCCATCTACAGGACCTT 984
Db      | 705 | TGCTCAAGAGAGTGGCTGCTTCTTCAACCTGAGTCCCTTACCCGCTGCTGAAAGCCCT 764
Oy      | 985 | CCTCAACCTGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1044
Db      | 765 | CCTCAACCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 824
Oy      | 1045 | AGCAGCAGCAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1104
Db      | 825 | AACGCAAGAGAGAGAGTCTGATGATGATGATGATGATGATGATGATGATGATGAT 884
Oy      | 1105 | CGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1164
Db      | 885 | AACGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 944
Oy      | 1165 | ACATAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1224
Db      | 945 | AGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1004
Oy      | 1225 | GGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1284
Db      | 1005 | GCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1064
Oy      | 1285 | GGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1344
Db      | 1065 | GATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1124
Oy      | 1345 | TGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1404
Db      | 1125 | TGGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1184
Oy      | 1405 | GTTGTGCTTACCTTACGAAAGGTTTCCAGCTTCAAGATGCTGCGGCTGAGGCTAT 1464
Db      | 1185 | GCTGTGCTTACCTTACGAAAGGTTTCCAGCTTCAAGATGCTGCGGCTGAGGCTAT 1244
Oy      | 1465 | ATTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1524
Db      | 1245 | ATTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1304
Oy      | 1525 | TTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1584
Db      | 1305 | TTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1364
Oy      | 1585 | CACCCCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1644
Db      | 1365 | --CCACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1388

```

```

US-10-007-262-2
; Sequence 2, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US-10/007,262
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-007-262-2
Query Match      38.5%; Score 741.6; DB 13; Length 2032;
Best Local Similarity 76.4%; Pred. No. 3,9e-228;
Matches 937; Conservative 0; Mismatches 288; Indels 5; Gaps 2;
Oy      | 388 | TCTCCTTCTCCGAGATGATGCTGTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAG 447
Db      | 165 | TCTTCACTTCCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 224
Oy      | 448 | CCCAGCTATGCTGTACTCTCTTCAATCCATATGCTGCTGCTGCTGCTGCTGCTG 506
Db      | 225 | CCCAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 284
Oy      | 507 | --AGGAGAGATCCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 564
Db      | 285 | CTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 344
Oy      | 565 | CCTCTTTTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 624
Db      | 345 | CTTCTTTTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 404
Oy      | 625 | CTTGCTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 684
Db      | 405 | CTTGCTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 464
Oy      | 685 | GGGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 744
Db      | 465 | GGGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 524
Oy      | 745 | CAGGCCCCGGGAACAGCAGCAGCTTCCAGTGGGAGCAAGCCGGGCGCTGCTGCTG 804
Db      | 525 | CTGGTCCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAGCAAGCCGGGCGCTGCTG 584
Oy      | 805 | CGCCTGTGTGTGATCTTCTTCCCTGCGCCAGAGATGCTCAGCCAGCAGTCAAGCTG 864
Db      | 585 | CACCTGCTGTGATCATCTCCACAGATGAAATCATCTCCCGGGGCTCAGTCAAGCTG 644
Oy      | 865 | TCTGCGGTGACAGCAGCCTTTGATATGCTGAGAAAGGCGCTGCTGCTGCTGCTG 924
Db      | 645 | TGTGAGTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 704
Oy      | 925 | TACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 984
Db      | 705 | TGTGAGTCAACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 764
Oy      | 985 | CCTCAACCTGACAGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1044
Db      | 765 | CCTCAACCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 824
Oy      | 1045 | AGCAGCAGCAGTATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1104
Db      | 825 | AACGCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 884

```



```

; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadamatsu, Kenji
; APPLICANT: Kadamatsu, Kenji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-
; TITLE OF INVENTION: SULFOTRANSFERASE AND
; TITLE OF INVENTION: DNA ENCODING THE SAME
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/10/212.933
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US/09/263.023
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: JP 10-54007
; PRIOR FILING DATE: 1998-03-05
; PRIOR APPLICATION NUMBER: JP 10-177844
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 3
; LENGTH: 2409
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (390)...(1841)
; US-10-212-933-3

```

```

Query Match      4.9%; Score 94; DB 14; Length 2409;
Best Local Similarity 50.7%; Pred. No. 6.4e-19;
Matches 288; Conservative 0; Mismatches 265; Indels 15; Gaps 2;

```

```

QY 535 TCGTGGTCTGCTCTTCTCGCGGTGACAGATCTTTTGGGACAGCTTTGCGGACG 594
DB 751 TGTACGTCTTCACACACGTGGCGCTCTGCTCTTCTCGCGGTATTTCAACGAGA 810
QY 595 ACCGGATGTTTCTACCTGATGAGCGCTGCGCATGTTGATGATGATGATGATGATG 870
DB 811 ATCCGAGGTGTTTCTCTACGAGCAGTGTGATGATGATGATGATGATGATGATGATG 870
QY 655 GCACAGCCTGGAAGCTGACATGCTGCGGATCTTCTGCGTCTTCTGCGTCTTCTGCGT 714
DB 871 GGGAGCGCGTTTCCCTGAGGCGGAGCGGAGCATGCTGAGCGCTTTACCGGTGCG 930
QY 715 ACATGAGCGTCTTGTGATGCTGATGATGATGATGATGATGATGATGATGATGATG 762
DB 931 ACCCTCTGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 990
QY 763 CCACGCTCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 822
DB 991 TGGGATCTTCTGCGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1050
QY 823 TCCCTGCCACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 879
DB 1051 ACCGAGAGAGGTGCTGGGCTGCTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1110
QY 880 CTTTGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 939
DB 1111 GCGTGGCGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1170
QY 940 GTTTCATGAGCTGAGGCTCTATGATGATGATGATGATGATGATGATGATGATGATG 999
DB 1171 GCGCTTCTGAGCTGCGGCTCTTGGCGCTCTGCTGCTGCTGCTGCTGCTGCTGCTG 1230
QY 1000 TCGTGAACCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1059
DB 1231 TCATCACTGCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1290
QY 1060 AACCTATGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1087
DB 1291 GCTCATCTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1318

```

RESULT 7

US-10-212-933-1
Sequence 1, Application US/10212933
Publication No. US2003008366A1

```

; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadamatsu, Kenji
; APPLICANT: Kadamatsu, Kenji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-
; TITLE OF INVENTION: SULFOTRANSFERASE AND
; TITLE OF INVENTION: DNA ENCODING THE SAME
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/10/212.933
; CURRENT FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US/09/263.023
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: JP 10-54007
; PRIOR FILING DATE: 1998-03-05
; PRIOR APPLICATION NUMBER: JP 10-177844
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO: 1
; LENGTH: 2150
; TYPE: DNA
; ORGANISM: Mus musculus
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (470)...(1918)
; US-10-212-933-1

```

```

Query Match      4.6%; Score 88.6; DB 14; Length 2150;
Best Local Similarity 50.1%; Pred. No. 3.3e-17;
Matches 285; Conservative 0; Mismatches 269; Indels 15; Gaps 2;

```

```

QY 534 GTGCTGCTGCTCTTCTCTGCGGTGACAGATCTTTTGGGACAGCTTTGCGGACG 593
DB 827 GTATATGCTTACACACGTGGCGCTCTGCTCTTCTGCGGTGATGATGATGATGATG 886
QY 594 CACCGGATGTTTCTACCTGATGAGCGCTGCGCATGTTGATGATGATGATGATGATG 653
DB 887 AACCTGAGGTGTTCTCTCTATGATGATGATGATGATGATGATGATGATGATGATG 946
QY 654 AGCAGAGCTGGAAGCTGACATGCTGCGGATCTTCTGCGTCTTCTGCGTCTTCTGCGT 713
DB 947 GGGAGCGCGTTTCCCTGAGGCGGAGCGGAGCATGCTGAGCGCTTTACCGGTGCG 1006
QY 714 GACATGAGCTCTTGTATGCTTACATGATGATGATGATGATGATGATGATGATGATG 764
DB 1007 GATCTTCTGCTTCTGATGATGATGATGATGATGATGATGATGATGATGATGATG 1066
QY 765 ---AGCCTCTTCAGTGGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 821
DB 1067 CTGGGACATCTTGGGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGC 1126
QY 822 TCCCTGCCACGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 878
DB 1127 TACGCGAAGAGAGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1186
QY 879 CCCTTTGATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 938
DB 1187 GCGCTGCGACCTTCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1246
QY 939 CGTTTTCACCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 998
DB 1247 CCGGCTTCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1306
QY 999 GTCTGACACTGTGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1058

```

[illegible]

Db 3 TAGTCATAAGGGTGTGGCGCTCTTCGACGTGGCGGCTTGGCGCCACTGCTGGAGACC 62

```

; SEQ ID NO 63
; LENGTH: 731
; TYPE: DNA
; ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc_feature
LOCATION: 237, 249, 263, 288, 312, 317, 323, 326, 337, 352, 362, 370,
LOCATION: 377, 409, 411, 414, 434, 436, 446, 457, 473, 486, 497, 498,
LOCATION: 502, 512, 531, 546, 554, 563, 565, 566, 588, 597, 608, 611,
LOCATION: 613, 615, 627, 632, 640, 641, 644, 654, 660, 663, 665
OTHER INFORMATION: n = A,T,C or G
FEATURE:
NAME/KEY: misc_feature
LOCATION: 671, 678, 692, 697, 698, 699, 704, 705, 712, 714, 717, 718,
LOCATION: 719, 723, 725, 730, 731
OTHER INFORMATION: n = A,T,C or G
US-10-117-982-63

Query Match      2.7%; Score 52.4; DB 12; Length 731;
Best Local Similarity 57.2%; Pred. No. 8.2e-06;
Matches 95; Conservative 0; Mismatches 71; Indels 0; Gaps 0;

OY          922 TGGTACATCAAGAGAGTGCGTTTTCACCCGACCGCCTCATCTACTACATCACAGACC 981
            ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB          3 TAGTCATTAAGGGGTGTGCCGCTTTTCAGCGTGGCGGCTTTGGCGGCACCTGCTGCAGACC 62

OY          982 CTTCOCCTAACCTGCACGCTGTGCACCTGTGTCCGAGACCCCCGGGCGCGTTCGCATCCC 1041
            ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB          63 CGGCCCTGCACCTCAAGAGTCAATCCCTGTGTCCGATGCCCGCGCGGCGAGTTCAC 122

OY          1042 GGGAGACACACCATTAAGATCATCANGTGAAGTCAAGTCAATATGTGCT 1087
            ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB          123 GGATTCGCTCGCGCCACGCGCTCATCCGCTGAGAGCTCAAGGTGCT 168

```


GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:31:05 ; Search time 124.227 Seconds
(Without alignments)
6843.163 Million cell updates/sec

Title: US-09-645-078-3

Perfect score: 1926
Sequence: 1 gggcatctactactactg.....ccgtgcacatctttcttaag 1926

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

- 1: Issued_Patents_NA:*
- 2: /cgn2_6/prodata1/1na/5A_COMB.seq:*
- 3: /cgn2_6/prodata1/1na/5B_COMB.seq:*
- 4: /cgn2_6/prodata1/1na/6A_COMB.seq:*
- 5: /cgn2_6/prodata1/1na/6B_COMB.seq:*
- 6: /cgn2_6/prodata1/1na/PCFUS_COMB.seq:*
- 7: /cgn2_6/prodata1/1na/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	741.6	38.5	2032	US-09-045-284A-1	Sequence 1, Appl1
2	741.6	38.5	2032	US-09-190-911-2	Sequence 2, Appl1
3	717	37.2	2065	US-09-786-240-26	Sequence 26, Appl1
4	114	5.9	2190	US-09-015-188-1	Sequence 1, Appl1
5	107.4	5.6	2156	US-08-899-514-1	Sequence 1, Appl1
6	98.4	5.1	2354	US-08-655-878-1	Sequence 1, Appl1
7	94	4.9	2409	US-09-263-023-3	Sequence 3, Appl1
8	94	4.9	2409	US-09-471-867-3	Sequence 3, Appl1
9	88.6	4.6	2150	US-09-263-023-1	Sequence 1, Appl1
10	88.6	4.6	2150	US-09-471-867-1	Sequence 1, Appl1
11	52.4	2.7	731	US-09-040-984-63	Sequence 63, Appl1
12	52.4	2.7	731	US-09-123-912-63	Sequence 63, Appl1
13	52.4	2.7	731	US-09-643-597-63	Sequence 63, Appl1
14	52.4	2.7	731	US-09-480-884A-63	Sequence 63, Appl1
15	52.4	2.7	731	US-09-542-615A-63	Sequence 63, Appl1
16	52.4	2.7	731	US-09-606-421B-63	Sequence 63, Appl1
17	45.2	2.3	7218	US-08-232-463-14	Sequence 14, Appl1
18	40.4	2.1	4352	US-09-620-312D-383	Sequence 383, Appl1
19	40	2.1	1233	US-09-266-965-24	Sequence 24, Appl1
20	40	2.1	1233	US-09-266-965-24	Sequence 24, Appl1
21	40	2.1	18331	US-09-266-965-96	Sequence 96, Appl1
22	39.8	2.1	1269	US-08-808-641-2	Sequence 2, Appl1
23	39.8	2.1	1269	US-09-064-839-2	Sequence 2, Appl1
24	39.8	2.1	1269	US-09-351-438-2	Sequence 2, Appl1
25	37.8	2.0	12597	US-09-705-299-12	Sequence 12, Appl1
26	37.6	2.0	1599	US-08-853-733B-1	Sequence 1, Appl1
27	37.6	2.0	1603	US-08-625-209A-1	Sequence 1, Appl1

28	37.6	2.0	1603	US-08-675-885-6	Sequence 6, Appl1
29	37.6	2.0	1847	US-08-675-885-4	Sequence 4, Appl1
30	37	1.9	1488	5268463-6	Patent No. 5268463
31	36.6	1.9	289	US-09-007-005-17	Sequence 17, Appl1
32	36.6	1.9	289	US-09-244-796-17	Sequence 17, Appl1
33	36.6	1.9	4057	US-08-894-997-49	Sequence 49, Appl1
34	36.4	1.9	1282	US-08-878-989-12	Sequence 12, Appl1
35	36.4	1.9	1282	US-09-272-796-12	Sequence 12, Appl1
36	36.4	1.9	1282	US-09-016-434-953	Sequence 953, Appl1
37	36.4	1.9	3223	US-07-980-528-1	Sequence 1, Appl1
38	35.4	1.8	990	US-09-252-991A-11517	Sequence 11517, A
39	35.4	1.8	2004	US-09-252-991A-11865	Sequence 11865, A
40	35.4	1.8	2166	US-09-252-991A-11722	Sequence 11722, A
41	35.2	1.8	1506	US-09-176-657-5	Sequence 5, Appl1
42	35.2	1.8	1506	US-09-421-299-5	Sequence 5, Appl1
43	35.2	1.8	1872	US-09-252-991A-9239	Sequence 9239, Ap
44	35	1.8	1287	US-09-252-991A-2964	Sequence 2964, Ap
45	35	1.8	1457	US-09-620-312D-408	Sequence 408, Ap

ALIGNMENTS

RESULT 1	US-09-045-284A-1	Sequence 1, Application US/09045284A
Patent No. 6265192		
GENERAL INFORMATION:		
APPLICANT:	Bistrup, Annette	
APPLICANT:	Rosen, Steven D.	
APPLICANT:	Hammerlich, Stefan	
TITLE OF INVENTION:	GLYCOSYL SULFOTRANSFERASE-3	
FILE REFERENCE:	6510-107051	
CURRENT APPLICATION NUMBER:	US/09/045,284A	
CURRENT FILING DATE:	1998-03-20	
NUMBER OF SEQ ID NOS:	9	
SOFTWARE:	FastSeq for Windows Version 4.0	
SEQ ID NO 1		
LENGTH:	2032	
TYPE:	DNA	
ORGANISM:	Homo sapiens	
US-09-045-284A-1		
Query Match	38.5%	Score 741.6; DB 3; Length 2032;
Best Local Similarity	76.4%	Pred. No. 2.7e-220;
Matches 937; Conservative	0; Mismatches 284; Indels	5; Gaps 2;
QY	388 TCTCTCTTCCGAGGATGATGCTTTGAAGAGGAGGAGGCTGATGCTCTGGGTT	447
DB	165 TCTTCCACTTCAGACAAATGCTACTGCTTAATAAATGAAGCTCTGCTGTTGTTT	224
QY	448 CCCAGTCATGCTGTTAGCTCTCTTCATCATATATGCTCCGACAGACACCTTCCAG-	506
DB	225 CCCAGATGAGGCAATCTTGGCTCTTATCTTCCATGATGACACCAACATAGCTCCCTGT	284
QY	507 --AGGAGAGAGTCAGAGAGCCGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	564
DB	285 CTATGAGAGGACACCCGACGCGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT	344
QY	565 CCTCTTTTGGGACAGCTTTTGGGAGGACCCGATGCTTCTTACCTGATGAGGCTG	624
DB	345 CTCTCTTTGGGGGACCTTTTGGGAGGACCCGATGCTTCTTACCTGATGAGGCTG	404
QY	625 CTTGGGATGTTGGATGATCTTACACGACGACGCTGAGAGCTGACATGCTGCTG	684
DB	405 CTTGGGATGTTGGATGATCTTACACGACGACGCTGAGAGCTGACATGCTGCTG	464
QY	685 GGGATCTTGGGATGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	744
DB	465 GGGATCTTGGGATGATGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG	524
QY	745 CAGGCCCGGAAACAGTCCAGCTTCTTCCAGTGGGAGGAGGAGGAGGAGGAGGAGGAG	804


```

Db 525 CTGTCCTCCGGAGACATGTCACCCCTCTTTCAGTGGGAGAACGCGGGCCCTGTGTTCG 584
Oy 805 CGCCTGTGTGACTTCTTCCCTGCGCCACGAGATAGCTCACCAAGCACTGACAGTGC 864
Db 585 CACCTGGCTGTGACTATCCACAGATGAATATATCCCGGGCTCACTGACAGCTCC 644
Oy 865 TCTGGGTGACAGCCCTTTTGTATATGTGAGAGAGCCCTGCGCTCTCAAGGCTTGTGG 924
Db 645 TGTGAGTCAACAGCCCTTTGAGTGTGAGAGAGCCCTGCGCTCTCAAGCAGCTGG 704
Oy 925 TACTCAAGAGAGTGGCTTTTTCAGCTGACAGGCGCTCTATCCATCTACTACAGGACCTT 984
Db 705 TGTCTAAGAGAGTGGCTTTTTCAGCTGACAGTCCCTCTACCGCTGCTGAAAGACCTT 764
Oy 985 CCCTCAACCTGACAGCTGTGACAGCTGTGCGAGACCCCGGGCGCTGTTCGATCCGAG 1044
Db 765 CCCTCAACCTGATATGCTGACAGCTGTGCGAGACCCCGGGCGCTGTTCGATCCGAG 824
Oy 1045 AGCAGACCAACCATAGACTGATGTTGACAGTCAATATGCTAGGCGACAGCTTTGGAAA 1104
Db 825 AACGCAAAAGGAGATCTCATGATTTGACAGTCCGATTTGATGGGCGAGCATGAGCAGA 884
Oy 1105 CGATCAAGAGAGAGAGAGAGAGCCCTATATAGCCATGAGATCATCTGCAAAAAGCAGTGG 1164
Db 885 AACTCAAGAGAGAGAGAGAGAGCCCTATATAGTATGATGACAGTCACTGCGCAAGCAGCTGG 944
Oy 1165 ACATAGTCAAGAGAGAGAGAGAGAGCCCTGAAAGCTGTGACAGAGCGTACTGTTCTGA 1224
Db 945 AGATCTAAGAGAGAGAGAGAGAGAGCCCTGCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1004
Oy 1225 GGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1284
Db 1005 GGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1064
Oy 1285 GGTGAGATTTTGTGCGCCAGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1344
Db 1065 GATGAGATTTTGTGCGCCAGCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1124
Oy 1345 TGGGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1404
Db 1125 TGGGTGACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1184
Oy 1405 GTTGTGCTTACCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1464
Db 1185 GCTGTGCTTGTGCGCCATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1244
Oy 1465 ATTGCTGGAGATCTCCAGAGTCAAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1524
Db 1245 ATTGCTGGAGATCTCCAGAGTCAAGATCTCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1304
Oy 1525 TTTGTGCTTGTGCGCCATGATCTTTGGGAGAGTCTTCCGAGAGAGAGAGAGAGAG 1584
Db 1305 TTTGTGCTTGTGCGCCATGATCTTTGGGAGAGTCTTCCGAGAGAGAGAGAGAGAG 1364
Oy 1585 CACCCCTGTGCTGACAGCTTGTAGTAC 1610
Db 1365 --CCACCTGTGTGACAGCTTGTAGTAC 1388

```

```

RESULT 2
US-09-190-911-2
; Sequence 2, Application us/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerlich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; EARLIER FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284

```

```

; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
; US-09-190-911-2

Query Match      38.5%; Score 741.6; DB 4; Length 2032;
Best Local Similarity 76.4%; Pred No. 2.7e-220;
Matches 937; Conservative 0; Mismatches 284; Indels 5; Gaps 2;

Oy 388 TCTCCTTCTCCGAGATGATGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 447
Db 165 TCTCCTCCTTCAAGCAATGCTAGTCTTAAAAAATGAGAGAGAGAGAGAGAGAGAGAG 224
Oy 448 CCCAGGTATGCTTGTAGCTCTCTTCAATATATGCTGAGAGAGAGAGAGAGAGAGAG 506
Db 225 CCCAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 284
Oy 507 --AGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 564
Db 285 CATTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 344
Oy 565 CCTTTTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 624
Db 345 CTTTCTTTGTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 404
Oy 625 CTTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 684
Db 405 CTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 464
Oy 685 GGGATCTTCTGCGTCTCTCTCTCTGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAG 744
Db 465 GGGATCTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 524
Oy 745 CAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 804
Db 525 CTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 584
Oy 805 GCGGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 864
Db 585 CACTGCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 644
Oy 865 TCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 924
Db 645 TGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 704
Oy 925 TACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 984
Db 705 TGTCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 764
Oy 985 CCCTCAACCTGACAGCTGTGACAGCTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 1044
Db 765 CCCTCAACCTGATATGCTGACAGCTGTGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 824
Oy 1045 AGCAGACCAACCATAGACTGATGTTGACAGTCAATATGCTGAGAGAGAGAGAGAGAG 1104
Db 825 AACGCAAAAGGAGATCTCATGATTTGACAGTCCGATTTGATGGGCGAGCATGAGCAGA 884
Oy 1105 CGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1164
Db 885 AACTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 944
Oy 1165 ACATAGTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1224
Db 945 AGATCTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1004
Oy 1225 GGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1284
Db 1005 GGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1064

```

QY	1285	GGTGGAAATTTTGGCCCACTCCAAACATGGGTTTACAAATGTACCCGGGGCAAGGGCA	1344
Db	1065	GATTGGAAATCTTGGCCCATCTTTAGACCTGGGTGCATTAACATCACCCGAGCAAGGGCA	1124
QY	1345	TGGGTGCAGCATGCTTCATCTACTAAGCAGAGAAAGCCCTCAAGCTCTCAGGCGTGGC	1404
Db	1125	TGGGTGCAGCAGCTTTTCCACACAAATGCGAAGGATGCCCTTAATGTCTCCAGGCTTGGC	1184
QY	1405	GTTGGTCTTACCTCTTACGAAAAGGTTTCCAGCTTCAAGATGCTCGCGGTAGGCTATGG	1464
Db	1185	GCTGGTCTTGGCCCTATGAAAAGTTTCTGCACCTTCAGAAAAGCCTGTGGCATGCATGA	1244
QY	1465	ATTGTGCGGATACCTCCAGGTCAGATCTCAACAAGAACAGGCAACCTGTCCCTGGATC	1524
Db	1245	ATTGTGCGGCTACCGCACGCTCAGATCTGAACAAAGAACAGAAACCTGTGTGGATC	1304
QY	1525	TTCGTCTCTCTCCCATATCTTTGGGGCAGGCTTTCGAGAAAGTTTAAGAGGCTGTCTG	1584
Db	1305	TTCGTCTACTCGACATGTCCTCCACGCAAAATCCACTAAGAGGGTTGAGAAAGCTTTCCTG	1364
QY	1585	CACCCCTTGGTTCAGGCTTAGTAC	1610
Db	1365	--CCACCGTGTGACGCTCAGTAC	1388

RESULT 3

US-09-786-240-26
Sequence 26 Application US/09786240
Patent No. 6558935
GENERAL INFORMATION:
APPLICANT: INCYTE PHARMACEUTICALS, INC.
APPLICANT: TANG, Y. Tom
APPLICANT: CORLEY, Neil C.
APPLICANT: GUEGLER, Karl J.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: LAL, Preeti
APPLICANT: YUE, Henry
APPLICANT: HILLMAN, Jennifer L.
APPLICANT: AZIMZAI, Yalda
TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
FILE REFERENCE: PF-0592 PCT
CURRENT APPLICATION NUMBER: US/09/786,240
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 60/133,642
PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1999-05-11
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PERL Program
SEQ ID NO 26
LENGTH: 2065
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc.feature
OTHER INFORMATION: Incyte ID NO. 6558935 2617407CBI
US-09-786-240-26

Query Match	37.2%	Score 717;	DB 4;	Length 2065;
Best Local Similarity	76.2%;	Pred. No. 1.2e-212;		
Matches 935;	Conservative 0;	Mismatches 285;	Indels 7;	Gaps 4;

QY	388	TCCTCTCTTCCGAGAGATATGCTGTTGAAGAAAGGAGCGTGTATGTTCTCGGGT	447
Db	157	TCCTTCACCTTAGACACATGCTACTGCGCTAAAAAAATGAAATCTCTGCTGTTTCGGTTT	216
QY	448	CCGAGCATGCTGTAGCTCTCTTCATCCATATGTCGTCACAGACACCTTCCAG-	506
Db	217	CCCGAATGCGCATTTGGCTCTATTCTTTCACATGTACAGCACAAATCAGCTCCCGT	276
QY	507	--AGGAGAGATCCAGAGAGCCCGCTGATGCTGCTCTCTCTCCGAGGCTCAGAT	564
Db	277	CTATGAGAGCACACCCGAGCGCATGCACGTGCTGTCTCTCTCTGCGGCTTGTGCT	336
QY	565	CCCTCTTTTGTGGACAGCTTTTCGGGGACACCCGGATGTGTTACCTGATGAGACCTG	624

Db	337	CTTCTTTTGTGGGACGCTTTTGTGGAGCACCCAGATGTTTCTTACTCATGTGAGGCCG	396
OY	625	CTTGGCATGTGTGATGACTTTTACCAGCAGCAGAGCTGGAAAGCTGCACATGGCTGTGC	684
Db	397	CTGTGCACAGTGTGGATGATACCTTTCAAGCAGACAGCCGCTGGATGCTGCACATGGCTGTGC	456
OY	685	GGGATCTTCTGCGTTCGCTCTTCTGTGTGACATGAGCGTCTTGGATGCTCATGTAAAC	744
Db	457	GGGATCTGTATACGGGGCGGTCTTCTTGTGGCAGATGACAGTCTTGGATGCTCATGTAAAC	516
OY	745	CAGGCCCCCCGGAAACATGCCAGCCTCTTCCAGTGGGAGCAAAAGCCGGGCCCTGTGCTCAG	804
Db	517	CTGTGTCCCGGAGACATCCAGCCCTTTTCAGTGGGAGAAACAGCCGGGCCCTGTGCTCTG	576
OY	805	CGCCTGTGTGTGATTTCTTCCCTCCACAGGATACACTCACCACCAACACTGTGCATGCTGC	864
Db	577	CACCTGCCCTGTGACATATATCCACAAAGATGA-ATCATCTCCCGGGGCTCATGTGCAGCTCC	635
OY	924	GTACTCAAGAGGTGCGCTTTTCTAGCCTGAGGCCCTTATCTACATCTACAGGACCT	983
Db	696	GTGTCTCAAGGAGGGGCGCTTCTTCAACCTGTGAGTCCCTTACCCGCTGTGAAAGACCC	755
OY	984	TTCCTCAACCGCAGCTGTGTGACACCTGTGTCGAGAACCCCGGGCGGTTCGATGCCGG	1043
Db	756	TCCCTCAACCTGCATATCTGTGACCTGTGTCGAGAACCCCGGGCGGTTCGATGCCGA	815
OY	1044	GAGCACACCAACCATAGAACTCATGTTGACAGTCAATTTGTCTGAGGGCAGCATTTTGGAA	1103
Db	816	GAAGCGACAAAGGAGAGATCTCATATTGACAGTGTGATTTGATGAGGGCAGCATAGACAG	875
OY	1104	ACGATCAAGAGGAAAGACAGCCCTATTATGCCATGAAAGTCAATCMTGCCAAAGCCAGCTG	1163
Db	876	AAACTCAAGAGGAGGAGACCAACCTCTACTATGTATGTAGAGGTCACTCTCCAAAGCCAGCTG	935
OY	1164	GACATAGTCAAGGCGCATATCCAAACCTCCCTGTGAAAGCTGTGAGCAGCGCTACTGTCTGTG	1223
Db	936	GAGATCTCAAGAGACATATCCATCTCTTGTGCCAAGGCCCTGCAAGAAAGCTTACCTGTGTG	995
OY	1224	AGGTATAGAGACCTGTGTTGGGGCACCCCTGGGCCAGACGACAGCATATATTAATTTGTG	1283
Db	996	CGCATATAGAGACCTGTGTTGGGGCACCCCTGGGCCAGACCTCCCAATATGATTAATTTGTG	1055
OY	1284	GGGTGTGATTTTGTGCCCCACCTCCCAACATGGGTTTCAATGTACCCCGGCGAAGGGC	1343
Db	1056	GGATTTGAAATCTTGTGCCCATCTTTCAGACCTGGGTGCAATACATACCCGAGGCAAGGGC	1115
OY	1344	ATGGGTACAGCATGCTCTTACTACTAACGCGCAGGAAAGCCCTCAACGTCTCTCAAGGCTGG	1403
Db	1116	ATGGGTACAGCATGCTCTTTCACACAAAGTCCAGAGGATGCCCTTATATGTCTCCAGAGCTTGG	1175
OY	1404	CGTTGTGCTTACCTTACGAAAGGTTTCCAGCTTCAAGATGCTCTGGGTGAGGCTATG	1463
Db	1176	CGTGTGTCTTGTGCCCCCTTGAAGAAAGTTTCTGCACTTGAAGAAAGCTGTGGCAGTGCATG	1235
OY	1464	GATTTGCTGGGATACCTCCAGGTAGATATTCACAAAGCAAGGCAAGCAAGCTGTCCCTGAT	1523
Db	1236	AATTTGCTGGGCTTACCCGACAGCTAGATCTTGAAACAAAGACAGAAACCTGTGTGAT	1295
OY	1524	CTTCTGTCTCTCCCATATCTTTGGGGCAGGTCTTTCGAGAAAGGTTAAGAGGCTGTCT	1583
Db	1296	CTTCTGTCTCTCCCATATCTTTGGGGCAGGTCTTTCGAGAAAGGTTAAGAGGCTTTGCT	1355
OY	1584	GCACCCCTTGGTTCCAGCCTTAGTCAC	1610
Db	1356	G--CCACCTGTGTCAAGCTTCAGTCAC	1380

RESULT 4

337 CTTCTTTTGGGGGAGCTTTTGGGAGACCCAGATGTTTCTTACCTGATGGAGCCG 396

625 CCTGGCATGTGGATGATGACTTTTACACGACGACAGCCTGGAAAGTCAGATATGGCTGTGC 684

397 CCTGGACAGTGGATGATGACCTTTCAGAGAGACAGCAGCCGTGATGCTGCACATGGCTGTGC 456

685 GGGATTCCTCTGCGTTCCGTTCTCTGTGTGACATGAGCCTTTTGATGCTACATGAAC 744

457 GGGATCTATACGGGCGCTCTTCTTGTGGACATGAGAGCTTTTATGCTTACATGMAAC 516

745 CAGGGCCCCGGAAACAGTCCAGCCCTTCCAGTGGGAGGAAGCCGGGCGCTGTGCACG 804

517 CTGGTCCCGGAGACAGTCCAGCCCTTTTGAAGGGAAGAACCGGGCGCTGTGCTCG 576

805 CGCCTGTGTGATCTTCTTCCCTGCCACAGATAGCTACACCAAGACATGCAAGCTGC 864

577 CACCTTGCTGTGATCTATCTCCACAGATATA-ATCATCCCCGGGCTCAGCTGACAGCTCC 635

865 TCTGGGCTCAGACGCCCCCTTG-ATATGTTGGAGAAAGGCTGCCGCTCTCACGGCTTG 923

636 TGTGAGTCAACAGCCCTTTGAATGTTGTGAAGAAAGCTGCGGCTCTCAAGCACAGC 695

924 GTACTTCAGAGAGTGCCTTTCTAGCCTCAGAGCCCTCTATCCACTACTACAGGACCT 983

696 GTGCTCAAGAGAGTGCCTTTCTTCAACCTCAGATCCCTTACCCGCTCTGTAAGAACCC 755

984 TCCCTCAACCTGCACAGTGTGCACCTGGTCCGAGACCCCCGGGCGGTGTCGATCCCG 1043

756 TCCCTCAACCTGCATATGTGTACACTGTGTCCGGAGACCCCCGGGCGGTGTCCGTA 815

1044 GAGCACACCACCATAGAACTCATGTGTGACAGTCAATATTTGTCTGAGGGACCATTTTGA 1103

816 GAACCCCAAGAGGAGATCTCATGTTTGAAGTCCGATTTGTATGGGACACATAGACAG 875

1104 AGCATCAAGAGAGAAAGACACACCCTATTATGCCATGAATCATCTGCAAAAGCAGGTG 1163

876 AAACCTCAAGAGAGAGACCAACCTCTACTATGATATCAGAGTATCTGCGCAAGCCAGCTG 935

1164 GACATATGTCAGAGCCATCCAAACCCTCCCTGGAAGCTGTGCAGACAGCTACTGTCTCTG 1223

936 GAGATCTCAAGAGACCATCAGTCACTCTTGTCCAAAGCCCTGCAGAAAGCTACTGCTGTG 995

1224 AGGATGAGAGACCTGTGTCCGGGACCCCTGGGCCCCAGACAGACATATTAATTAATTGTG 1283

996 CCGTATGAGAGACCTGTGTCCGGGACCCCTGTGTGGCCCAACTTCCGATATGAATTAATG 1055

1284 GGGTTGGAATTTTGGCCCCCAGCTTCAAAAGATGGGTTTAAATGTACCCGGGCAAGGGC 1343

1056 GGAATGGAATTTTGGCCCCCAGCTTCTCAACCTGGGTCATTAATCATCCCGAGGCAAGGC 1115

1344 ATGGGTCAGCATGCTTTCATACTTAAGGCCAGAGAACGCCCTCAAGCTCTTCAGGCGTGG 1403

1116 ATGGGTCAGCATGCTTTCATACTTAAGGCCAGAGAACGCCCTTAATGTCTCCAGAGCTTGG 1175

1404 CGTGGTCTTACCTTACGAAAGGCTTCCACCTTCAAGATATGCTCGGGGTAGGCTATG 1463

1176 CCGTGGTCTTACCTTACGAAAGGCTTCTCGACTTCAAGAAAGCCGTGTGGGATGCTATG 1235

1464 GATTTGCTGGGTCATCTCAGAGTCAATGTCACACAAAGAACAGGCAACCTGCTCCCTGGAT 1523

1236 AATTTGCTGGGTCATCTCAGAGTCAATGTCACACAAAGAACAGGCAACCTGCTCCCTGGAT 1295

1524 CTTTGTCTCTCTCCCATATCTTGGGGGAGGTCTTCCGAAAGGTTTAAAGAGTCTGTCT 1583

1296 CTTTGTCTCTCTCCCATATCTTGGGGGAGGTCTTCCGAAAGGTTTAAAGAGTCTGTCT 1355

1584 GCACCCCTTGGTTCAGGCTTAAGTCAC 1610

1356 G--CCACCTGTGTGTCACCTTAAGTCAC 1380

US-09-015-188-1
 ; Sequence 1, Application US/09015188C
 ; Patent No. 6393358
 ; GENERAL INFORMATION:
 ; APPLICANT: Williams, Kevin J
 ; APPLICANT: Tabas, Ira
 ; TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
 ; TITLE OF INVENTION: 6-Sulfotransferase
 ; FILE REFERENCE: JEP-0231
 ; CURRENT APPLICATION NUMBER: US/09/015,188C
 ; CURRENT FILING DATE: 1998-01-29
 ; NUMBER OF SEQ ID NOS: 17
 ; SOFTWARE: PatentIn Ver. 2.0
 ; SEQ ID NO 1
 ; LENGTH: 2190
 ; TYPE: DNA
 ; ORGANISM: Homo sapiens
 ; S-09-015-188-1
 Query Match 5.9%; Score 114; DB 4; Length 2190;
 Best Local Similarity 48.4%; Pred. No. 7, 4e-25;
 Matches 500; Conservative 0; Mismatches 490; Indels 42; Gaps 5;
 QY 531 CATGCTGCTGCTCTCTCTCTCTGCGGTCAGGATCCTCTTTGTGGACAGCTTTTCGGG 590
 DB 308 CACATCTCTATCTCTGCGCCACACGCGGAGGCTCTCTCTCTGCTGGCCAGCTTTCAAC 367
 QY 591 CAGCACCGGATGTGTCTTACCTGATGAGGCTGCTGCGATGTGTGATGATCTTCAAC 650
 DB 368 CAGCAGCTGAGAGCTGTCTTACCTGATGAGGCTGCTGCGATGTGTGATGATCTTCAAC 427
 QY 651 AGCAGCAGAGCTGAGAGTGCATGCTGTGCGG----- 686
 DB 428 CCCGCTTACACCGAGGAGGAGCCGCGGAGGCTGATGCTAGGCGCCAGCCG 487
 QY 687 GATCTTCTGCTCTCTCTCTCTCTCTGTCAGAGCTGCTTGTGATGCTTCAACATGAGCCCA 746
 DB 488 GACCTCTGCGGAGGAGCTCTAGAGCTGCGAGCTCTCTCTGAGAGCTTCAACATGAGCCG 547
 QY 747 GGGCCCCGGAAGAGCTCCAGCC-----TCTTCAAGTGGAGCAAAAGCGGCGCTGTGC 800
 DB 548 CCGCGGCTACACCGAGAGGAGGAGCTTCTGCGGCGGCGGAGGCGGCTCTCTGC 607
 QY 801 TCAGGCGCTGTGTGAG--TCTTCTCTGCGGAGAGAGTCAAGCTCAACCCAGCAGCTGC 857
 DB 608 TCCGCGCTGTGTGAGCTCTGCGGCGGAGGAGCTGTCTCTGAGAGGAGGAGGAGCTGT 667
 QY 858 AAGCTGCTCTGCGGTCAGAGAGCTTGTGATGTGTGAGAGAGGCTGCGCTCTCAAGCGC 917
 DB 668 GTGGCGAAGTGGGGCTACTCAAGCTGAGCGTGGGGCGGAGGCGTGGCGAGCGGAGC 727
 QY 918 TTCTGTGACTCAAGAGGTGCTTGTCTCAAGCTGCGAGCGCTGTATCTCACTACTCAGC 977
 DB 728 CACGTGGGCTATCAAGAGGTGCTGCGGCGGAGGAGTCAAGCTGCGGCGCTGTGTGMA 787
 QY 978 GACCTTCTCTCAAGCTGAGCTGCTGAGCTGCTGCGAGAGAGAGAGAGAGAGAGAGAGAG 1037
 DB 788 GACCTGCTGTAAAGCTCAAGGTCTCAAGGTCTGCGAGAGAGAGAGAGAGAGAGAGAG 847
 QY 1038 TCCCGGAG 1097
 DB 848 TCGCGAG 906
 QY 1098 TTGGAAGAGATCAAG 1157
 DB 907 GAGGAAGAGCT-----ACAACTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 958
 QY 1158 CAGGTGAGATGATGAG 1217
 DB 959 TTCTCAAGCTCTGCTCAAG 1018
 QY 1218 TTCTGAGATGAG 1277

DB 1019 TTGCTGCTCTACGAG 1078
 QY 1278 TTGTGGGGGTGATTTTGTGCGCCACCTCTCAAGAGAGAGAGAGAGAGAGAGAGAGAG 1337
 DB 1079 TTCTGAG 1138
 QY 1338 AAGGAGATGGGTGAG 1397
 DB 1139 GACCCAG 1198
 QY 1398 GCGTGGAG 1457
 DB 1199 AAGTGGAG 1258
 QY 1458 GCTATGATTTGCTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1517
 DB 1259 GTGCTGAG 1318
 QY 1518 CTGAGATCTCTG 1529
 DB 1319 GTGAGCTGTG 1330
 RESULT 5
 US-08-899-514-1
 ; Sequence 1, Application US/08899514
 ; Patent No. 5910581
 ; GENERAL INFORMATION:
 ; APPLICANT: HABUCHI, OSAMI
 ; APPLICANT: FUKUTA, MASAKAZU
 ; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
 ; TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
 ; NUMBER OF SEQUENCES: 9
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
 ; STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
 ; CITY: NEWPORT BEACH
 ; STATE: CALIFORNIA
 ; COUNTRY: US
 ; ZIP: 92660
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/899,514
 ; FILING DATE:
 ; CLASSIFICATION: 435
 ; PRIORITY APPLICATION DATA:
 ; APPLICATION NUMBER:
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: DANIEL E. ALTMAN
 ; REGISTRATION NUMBER: 34,115
 ; REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 714 760 0404
 ; TELEFAX: 714 760 9502
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 2156
 ; TYPE: nucleic acid
 ; STRANDEDNESS: both
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: cDNA
 ; ORIGINAL SOURCE:
 ; ORGANISM: Human
 ; TISSUE TYPE: Fetal brain
 ; FEATURE:
 ; NAME/KEY: CDS
 ; LOCATION: 147..1583

Db 1500 GGCTACAACTGG 1512

RESULT 6
US-08-655-878-1
; Sequence 1, Application US/08655878

```

GENERAL INFORMATION:
APPLICANT: FUKUTA, MASAKAZU
APPLICANT: HABUCHI, OSAMI
TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
NUMBER OF SEQUENCES: 9
CORRESPONDENCE ADDRESS:
ADDRESSEE:
STREET:
CITY:
STATE:
COUNTRY:
ZIP:
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/655,878
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE:
TELEFAX:
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2354
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA to mRNA
ORIGINAL SOURCE:
ORGANISM: Chick
TISSUE TYPE: Embryo chondrocyte
FEATURE:
FEATURE: NAME/KEY: CDS
FEATURE: LOCATION: 211..1584
FEATURE: IDENTIFICATION METHOD: P
FEATURE:
FEATURE: NAME/KEY: s4g_peptide
FEATURE: LOCATION: 211..309
FEATURE: IDENTIFICATION METHOD: P
FEATURE:
FEATURE: NAME/KEY: mat_peptide
FEATURE: LOCATION: 310..1584
FEATURE: IDENTIFICATION METHOD: P
FEATURE:
FEATURE: NAME/KEY: transmembrane domain
FEATURE: LOCATION: 280..321
FEATURE: IDENTIFICATION METHOD: P
FEATURE:
FEATURE: NAME/KEY: potential N-glycosylation site
FEATURE: LOCATION: 394..402
FEATURE: IDENTIFICATION METHOD: S
FEATURE:
FEATURE: NAME/KEY: potential N-glycosylation site
FEATURE: LOCATION: 427..435
FEATURE: IDENTIFICATION METHOD: S

```


Db 1051 ACCGCAAGAGGTGCTGGGGTGGTGACGACCCGCTGTGCAGAAAGAGCCCGCACAGC 1110
QY 880 CCTTGAATAGTGGAGAGAGCCCTGCGCTCTACAGGCTTGTGTACTCAAGAGGTGC 939
Db 1111 GCGTGGGCGCTTGTGAGAGAGAGTGCAGCACTGATCTAAGAGGTGTC 1170
QY 940 GTTTCACAGCTGACAGCCCTCTATCCACTACTACAGGACCTTCCCTCAACCTGCAGC 999
Db 1171 GCGTCTTCGACGTGGCGGTCTTGGCGCACCTGCTGCAGAGACCCGCGCTTGACCTCAAGG 1230
QY 1000 TCGTGACCTGTGTCGAGACCCCGGGCGGTTCGATCCGATCCGGGAGACACACACATAG 1059
Db 1231 TCATCCACTGTGTGCTGATCCCGCGGGTGGCGAGTTCAAGGATCCGCTCGCGCCACG 1290
QY 1060 AACCATGCTGACAGCATATGTGCT 1087
Db 1291 GCGTATCCGTGAGAGCCTACAGGTGT 1318

RESULT 8
US-09-471-867-3

Sequence 3, Application US/09471867
Patent No. 6455289

GENERAL INFORMATION:

APPLICANT: Uchimura, Kenji
APPLICANT: Muramatsu, Hideki
APPLICANT: Kadomatsu, Kenji
APPLICANT: Kanagaki, Reiji
APPLICANT: Habuchi, Osami

APPLICANT: Muramatsu, Takashi

TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
FILE REFERENCE: TOYAMA1.001AUS

CURRENT APPLICATION NUMBER: US/09/471,867

EARLIER FILING DATE: 1999-12-23

EARLIER APPLICATION NUMBER: US 09/263,023

EARLIER FILING DATE: 1999-03-05

EARLIER APPLICATION NUMBER: JP 10-54007

EARLIER FILING DATE: 1998-03-05

EARLIER APPLICATION NUMBER: JP 10-177844

EARLIER FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 10

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 3

LENGTH: 2409

TYPE: DNA

ORGANISM: Homo sapiens

FEATURE:

NAME/KEY: CDS

LOCATION: (390)...(1841)

US-09-471-867-3

Query Match 4.9%; Score 94; DB 4; Length 2409;
Best Local Similarity 50.7%; Pred. No. 1.3e-18;
Matches 288; Conservative 0; Mismatches 265; Indels 15; Gaps 2;

QY 535 TCGTGGTCTCTCTTCTGCGGGTCAAGATCTCTTTTGGGACAGCTTTTCGGGACG 594
Db 751 TGTACGTCTTACACAGCGTGGCGCTCTGCTCTTCTTGGGAGCTATTAACACGA 810
QY 595 ACCCGAGGTGTTTACCTATGAGACCTGCTGGAGTGTGATGATGATCTTCAACACGA 654
Db 811 ATCCGAGGTGTTTCTTCTTACAGAGCGAGTGTGATGATGATGATGATGATGATGATG 870
QY 655 GACAGAGCTGAGACATGAGCTGTGCGGATCTTCTGCTTCTTCTTCTTCTTCTTCTG 714
Db 871 GGGAGCGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTG 930
QY 715 ACATGAGCGCTTGTGATGCTTACATGAACCCAGGCCCCCGG-----AAACAGT 762
Db 931 ACCTCTCTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 990

QY 763 CCAGCTCTTCCAGTGGAGCAAGCCGCGCTGTGTCTACAGCCCTGTGTGTACTTCT 822
Db 991 TGGGATCTTCTTGGGAG 1050
QY 823 TCCCTGCGGAGAGATGAGCTTCAACCAAGACCTCAAGCT---GCTTGGGTGAGCAGC 879
Db 1051 ACCGCAAGAGAGTGTGGGGTGGTGACGACCCGCTGTGCAGAAAGAGCCCGCACAGC 1110
QY 880 CCTTGAATAGTGGAGAGAGCCCTGCTCTACAGGCTTGTGTGTACTCAAGAGGTGC 939
Db 1111 GCGTGGGCGCTTGTGAGAGAGAGTGCAGCACTGATCTAAGAGGTGTC 1170
QY 940 GTTTCACAGCTGACAGCCCTCTATCCACTACTACAGGACCTTCCCTCAACCTGCAGC 999
Db 1171 GCGTCTTCGACGTGGCGGTCTTGGCGCACCTGCTGCAGAGACCCGCGCTTGACCTCAAGG 1230
QY 1000 TCGTGACCTGTGTCGAGACCCCGGGCGGTTCGATCCGATCCGGGAGACACACATAG 1059
Db 1231 TCATCCACTGTGTGCTGATCCCGCGGGTGGCGAGTTCAAGGATCCGCTCGCGCCACG 1290
QY 1060 AACCATGCTGACAGCATATGTGCT 1087
Db 1291 GCGTATCCGTGAGAGCCTACAGGTGT 1318

RESULT 9
US-09-263-023-1

Sequence 1, Application US/09263023
Patent No. 6037159

GENERAL INFORMATION:

APPLICANT: Uchimura, Kenji
APPLICANT: Muramatsu, Hideki
APPLICANT: Kadomatsu, Kenji
APPLICANT: Kanagaki, Reiji
APPLICANT: Habuchi, Osami

APPLICANT: Muramatsu, Takashi

TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
FILE REFERENCE: TOYAMA1.001AUS

CURRENT APPLICATION NUMBER: US/09/263,023

EARLIER FILING DATE: 1999-03-05

EARLIER APPLICATION NUMBER: JP 10-54007

EARLIER FILING DATE: 1998-03-05

EARLIER APPLICATION NUMBER: JP 10-177844

EARLIER FILING DATE: 1998-06-24

NUMBER OF SEQ ID NOS: 10

SOFTWARE: FastSeq for Windows Version 3.0

SEQ ID NO 1

LENGTH: 2150

TYPE: DNA

ORGANISM: Mus musculus

FEATURE:

NAME/KEY: CDS

LOCATION: (470)...(1918)

US-09-263-023-1

Query Match 4.6%; Score 88.6; DB 3; Length 2150;
Best Local Similarity 50.1%; Pred. No. 5.9e-17;
Matches 285; Conservative 0; Mismatches 269; Indels 15; Gaps 2;

QY 534 GTGCTGGTGTCTCTCTTCTGCGGAGAGATCTTGTGGGACAGCTTTTCGGGAG 593
Db 827 GTGATGTGTTTACACAGCTGGCGCTCGGCTGTCTCTTCTTCTTCTTCTTCTTCTTCT 886
QY 594 CACCGAGTGTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 653
Db 887 AACCTGAGGTGTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCT 946
QY 654 AGCAGAGCTTGAAGCTGACATGAGCTGTGCGGATCTTCTGCTTCTTCTTCTTCTG 713
Db 947 GGGAGCGGCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTTCTG 1006
QY 714 GACATGAGCGTCTTGTATGCTTACATGAACCCAGGCCCCCGGAAACAGTCC----- 764

Db 1007 GATCTTCGATTTTCAGCTGTATACCCCGCAGAGAGTGGGGGGGCAACCTCACCAGT 1066
QY 765 ---AGCCTCTTCAGTGGAGCAAGCCGGCCCTGTGCTCAGCGCTGTGTGACTTC 821
Db 1067 CTGGGATCTTTTGGGAGCAGCAGCTAACAAAGTGTGTGTCTCTGCGACTCTGTCTGCC 1126
QY 822 TTCCCTGCCAGAGATCAGCTCACCAGCAGCTGCAAGCTGCTC---TGCGGTACAGAG 878
Db 1127 TACCGCAAGAGAGTCTCGAGTGTGTGAGACAGCCGGGTGTGCAAAAAGTCCCACTCAA 1186
QY 879 CCCTTGATATGTGTGAGAGAGGCTCCGCTCTCAGAGCTGTGTGTGTACTCAAGAGGTG 938
Db 1187 CGCTTGACAGCTTCGAGAGAGAGTGTCCAGATACCGAGCGAGGTGTATCAAGGGCGTG 1246
QY 939 CGTTTCTGAGCCTGAGAGCCCTCTATCCACTACTCAGGACCTTCCTCCATCAGCTGAC 998
Db 1247 CGGGCTTTGATGTGTGCTGTGTGGGGCCGCTCTTAAGATCAGACCTTTGGACCTCAAG 1306
QY 999 GTGCTGACCTGTGTCCGAGAGCCCGGGCCGTGTCCGATCCGGAGACACACCACATA 1058
Db 1307 GTGATCCACTAGTATAGTATCTCTGTGTGTGCGAGCTCCCGCATCCGCTCGCTCAC 1366
QY 1059 GAACCTATGTTGACAGTATATTTGCT 1087
Db 1367 GGCCTATCCGGGAAAGCCTACAGGTGT 1395

RESULT 10
US-09-471-867-1
Sequence 1, Application US/09471867
Patent No. 6455289

GENERAL INFORMATION:
APPLICANT: Uchimura, Kenji
APPLICANT: Muramatsu, Hideki
APPLICANT: Kadamatsu, Kenji
APPLICANT: Kannagi, Reiji
APPLICANT: Habuchi, Osami
APPLICANT: Muramatsu, Takashi
TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
FILE REFERENCE: TOYAMA41.001AUS
CURRENT APPLICATION NUMBER: US/09/471,867
CURRENT FILING DATE: 1999-12-23
EARLIER APPLICATION NUMBER: US 09/263,023
EARLIER FILING DATE: 1999-03-05
EARLIER APPLICATION NUMBER: JP 10-54007
EARLIER FILING DATE: 1998-03-05
EARLIER APPLICATION NUMBER: JP 10-177844
EARLIER FILING DATE: 1998-06-24
NUMBER OF SEQ ID NOS: 10
SOFTWARE: FASTSEQ for Windows Version 3.0
SEQ ID NO 1
LENGTH: 2150
TYPE: DNA
ORGANISM: Mus musculus
FEATURE:
NAME/KEY: CDS
LOCATION: (470)...(1918)
US-09-471-867-1

Query Match 4.6%; Score 88.6; DB 4; Length 2150;
Best Local Similarity 50.1%; Pred. No. 5.9e-17;
Matches 285; Conservative 0; Mismatches 269; Indels 15; Gaps 2;

QY 534 GTGCTGCTGCTGCTCTCTCTGAGGAGATCTCTTTGTGGACAGCTTTCCGGGAG 593
Db 827 GTGTATGTGTACACAGGCGGCTGGGCTGCTCTCTCTGCGTATAGCTCTTCAACAG 886
QY 594 CACCCGAGTGTGTCTACCTGATGAGCCTGCTGCGATGTGTGATGACTTTCACAG 653
Db 887 AACCTGAGGTGTCTCTCTCTATGAGCCTGTGTGCGAGTGTGGCAAAAGTGTACCC 946

QY 654 AGCAGCCTGAGAGCTGACATGAGCTGTGCGGAGTCTTTCGCTTCCGTCTGTGT 713
Db 947 GGGAGACCGCTTCCCTGAGGGGAGAGCGCGGAGCATGTGTGAGCCTCTCTACCGCTGC 1006
QY 714 GATATAGGCTCTTTATGTCTCTATGATGAACCCAGGCCCCCGGAAAGATCC----- 764
Db 1007 GATCTTCGATTTTCAGCTGTATACCCCGCAGAGAGTGGGGGGGCAACCTCACCAGT 1066
QY 765 ---AGCCTCTTCAGTGGAGCAAGCCGGCCCTGTGCTCAGCGCTGTGTGACTTC 821
Db 1067 CTGGGATCTTTTGGGAGCAGCAGCTAACAAAGTGTGTGTCTCTGCGACTCTGTCTGCC 1126
QY 822 TTCCCTGCCAGAGATCAGCTCACCAGCAGCTGCAAGCTGCTC---TGCGGTACAGAG 878
Db 1127 TACCGCAAGAGAGTCTCGAGTGTGTGAGACAGCCGGGTGTGCAAAAAGTCCCACTCAA 1186
QY 879 CCCTTGATATGTGTGAGAGAGGCTCCGCTCTCAGAGCTGTGTGTGTGTACTCAAGAGGTG 938
Db 1187 CGCTTGACAGCTTCGAGAGAGAGTGTCCAGATACCGAGCGAGGTGTATCAAGGGCGTG 1246
QY 939 CGTTTCTGAGCCTGAGAGCCCTCTATCCACTACTCAGGACCTTCCTCCATCAGCTGAC 998
Db 1247 CGGGCTTTGATGTGTGCTGTGTGGGGCCGCTCTTAAGATCAGACCTTTGGACCTCAAG 1306
QY 999 GTGCTGACCTGTGTCCGAGAGCCCGGGCCGTGTCCGATCCGGAGACACACCACATA 1058
Db 1307 GTGATCCACTAGTATAGTATCTCTGTGTGTGCGAGCTCCCGCATCCGCTCGCTCAC 1366
QY 1059 GAACCTATGTTGACAGTATATTTGCT 1087
Db 1367 GGCCTATCCGGGAAAGCCTACAGGTGT 1395

RESULT 11
US-09-040-984-63
Sequence 63, Application US/09040984
Patent No. 6210883

GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Wang, Tongtong
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS
TITLE OF INVENTION: OF LUNG CANCER
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,984
FILING DATE: 18-MAR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE/DOCKET NUMBER: 210121.456
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-282-6031
TELEX:
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 731 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear


```
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (613)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (615)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (627)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (632)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (640)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (641)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (644)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (654)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (660)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (663)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (665)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (671)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (678)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (692)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (697)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (698)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (699)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (704)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (705)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (712)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (714)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (717)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (718)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (719)
OTHER INFORMATION: Where n is a, c, g or t
```

```
NAME/KEY: modified_base
LOCATION: (723)
Query Match
Best Local Similarity 57.2%; Score 52.4; DB 4; Length 731;
Matches 95; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 922 TGTACTCAAGAGGTGCTTTCTCAGCCTGACGCGCTCTATCTCACTACTACAGGACC 981
DB 3 TAGTCATAAAGGTGTGCGCGCTCTTCGACGTGGCGGTCTTGCGCACGTCTGCGAGACC 62
QY 982 CTTCCTCACTGACGCTGTCGACCTGTGCGAGACCCCGGCGCGGTTCCTCATCC 1041
DB 63 CGGCCCTGACCTCAAGGTCATCCACTGCTGCTGATCCCGCGCGGTGCGAGTTCAC 122
QY 1042 GGGACACACACCACTAGACATCATGTGACAGTCATATGTGCT 1087
DB 123 GGATCCGCTCGCGCCACGCGCTCATCCGTGAGAGCTACAGTGTGT 168
RESULT 13
US-09-643-597-63
Sequence 63, Application US/09643597
Patent No. 6426072
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaltanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aljun
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
FILE REFERENCE: 210121.45311
CURRENT APPLICATION NUMBER: US/09/643,597
NUMBER OF SEQ ID NOS: 369
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 63
LENGTH: 731
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(731)
OTHER INFORMATION: n = A,T,C or G
US-09-643-597-63
Query Match
Best Local Similarity 57.2%; Score 52.4; DB 4; Length 731;
Matches 95; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 922 TGTACTCAAGAGGTGCTTTCTCAGCCTGACGCGCTCTATCTCACTACTACAGGACC 981
DB 3 TAGTCATAAAGGTGTGCGCGCTCTTCGACGTGGCGGTCTTGCGCACGTCTGCGAGACC 62
QY 982 CTTCCTCACTGACGCTGTCGACCTGTGCGAGACCCCGGCGCGGTTCCTCATCC 1041
DB 63 CGGCCCTGACCTCAAGGTCATCCACTGCTGCTGATCCCGCGCGGTGCGAGTTCAC 122
QY 1042 GGGACACACACCACTAGACATCATGTGACAGTCATATGTGCT 1087
DB 123 GGATCCGCTCGCGCCACGCGCTCATCCGTGAGAGCTACAGTGTGT 168
RESULT 14
US-09-480-884A-63
Sequence 63, Application US/09480884A
```

Search completed: August 14, 2003, 17:44:32
Job time : 128.227 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:40:04 ; Search time 34.5891 Seconds
(without alignments)
2894.677 Million cell updates/sec

Title: US-09-645-078-4
Perfect score: 2042
Sequence: 1 MMLKGRLLMFLGSGVIVV.....NLSDLLSSHIILQGVFREG 388

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 830525 seqs, 258052604 residues
Total number of hits satisfying chosen parameters: 830525

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

SPTREMBL_23:*

- 1: sp_archaea:*
- 2: sp_bacteria:*
- 3: sp_fungi:*
- 4: sp_human:*
- 5: sp_invertebrate:*
- 6: sp_mammal:*
- 7: sp_mhc:*
- 8: sp_organelle:*
- 9: sp_phage:*
- 10: sp_plant:*
- 11: sp_protent:*
- 12: sp_virus:*
- 13: sp_vertebrate:*
- 14: sp_unclassified:*
- 15: sp_virus:*
- 16: sp_bacteriaph:*
- 17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	2042	100.0	388	11 Q9WUE5	Q9WUE5 mus musculu
2	2033	99.6	388	11 Q9RII1	Q9RII1 mus musculu
3	1513.5	74.1	386	4 Q9Y5R3	Q9Y5R3 homo sapien
4	1510.5	74.0	386	4 Q8NCG5	Q8NCG5 homo sapien
5	1470.5	72.0	370	4 Q8IV46	Q8IV46 homo sapien
6	984	48.2	395	11 Q9OUP4	Q9OUP4 mus musculu
7	977	47.8	395	4 Q8GZK3	Q8GZK3 homo sapien
8	958.5	46.9	411	4 Q8GZS8	Q8GZS8 homo sapien
9	595.5	29.2	484	11 Q9EP78	Q9EP78 mus musculu
10	593	29.0	486	4 Q75667	Q75667 homo sapien
11	592.5	29.0	484	11 Q99NB0	Q99NB0 mus musculu
12	592	29.0	486	4 Q9NSR4	Q9NSR4 homo sapien
13	579.5	28.4	530	11 Q88276	Q88276 mus musculu
14	576.5	28.2	483	4 Q9UED5	Q9UED5 homo sapien
15	576.5	28.2	530	4 Q9Y4C5	Q9Y4C5 homo sapien
16	527.5	25.8	411	4 Q43916	Q43916 homo sapien

17	527.5	25.8	411	11 Q9EQC0	Q9EQC0 mus musculu
18	497	24.3	479	4 Q75099	Q75099 homo sapien
19	478.5	23.4	472	11 Q88199	Q88199 mus musculu
20	454.5	22.3	474	11 Q9QZL2	Q9QZL2 rattus norv
21	399	19.5	441	13 Q93403	Q93403 torpedo cal
22	281	13.8	486	5 Q9VMC3	Q9VMC3 drosophila
23	210.5	10.3	363	5 Q9VMC4	Q9VMC4 drosophila
24	203	9.9	120	11 Q9MDK5	Q9MDK5 mus musculu
25	199.5	9.8	363	5 Q8MZD1	Q8MZD1 drosophila
26	174	8.5	119	6 Q95J48	Q95J48 oryctolagus
27	128	6.3	274	16 Q8PRA0	Q8PRA0 xanthomomas
28	126	6.2	1222	4 Q8IZU8	Q8IZU8 homo sapien
29	124	6.1	307	16 Q92VC4	Q92VC4 rhizobium m
30	98.5	4.8	3187	11 Q63714	Q63714 rattus norv
31	98	4.8	913	2 Q9K1D4	Q9K1D4 streptomyce
32	97	4.8	303	16 Q93JF6	Q93JF6 streptomyce
33	96.5	4.7	379	3 Q8N1Z3	Q8N1Z3 neurospora
34	96.5	4.7	834	5 Q9U700	Q9U700 eutrochium
35	96.5	4.7	985	2 P94324	P94324 bradyrhizob
36	96	4.7	876	12 Q8AYX3	Q8AYX3 porcine lym
37	95	4.7	285	2 Q8KRX0	Q8KRX0 xanthomomas
38	95	4.7	388	2 Q87614	Q87614 pseudomonas
39	94.5	4.6	506	8 Q8MCS2	Q8MCS2 galega orle
40	94.5	4.6	513	16 Q9LAX8	Q9LAX8 escherichia
41	93.5	4.6	522	16 Q8CXZ3	Q8CXZ3 escherichia
42	93	4.6	346	10 Q9FZ80	Q9FZ80 arabidopsis
43	93	4.6	726	16 Q9X839	Q9X839 streptomyce
44	93	4.6	865	12 Q913V4	Q913V4 porcine gam
45	93	4.6	876	12 Q8JYD3	Q8JYD3 porcine lym

ALIGNMENTS

RESULT 1

ID	Q9WUE5	PRELIMINARY:	PRT:	388 AA.
AC	Q9WUE5:			
DT	01-NOV-1999 (TREMBLrel. 12, Created)			
DT	01-MAR-2003 (TREMBLrel. 23, Last sequence update)			
DE	N-acetylglucosamine 6-O-sulfotransferase (Carbohydrate (Chondroitin 6/keratan) sulfotransferase 4).			
GN	CHST4.			
OS	Mus musculus (Mouse).			
OC	Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.			
OX	NCBI_TaxID=10090;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Tongue;			
RC	MEDLINE=99264336; PubMed=10330415;			
RA	Huang C.-C., Kannagi R., Rosen S.D., Hemmerich S.,			
RA	"Sulfotransferases of two specificities function in the reconstitution			
RT	of high endothelial cell ligands for L-selectin.";			
RL	J Cell Biol. 145:899-910(1999).			
RL	[2]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=C57BL/6J; TISSUE=Tongue;			
RC	MEDLINE=21085660; PubMed=11217851;			
RA	Kawai T., Shinagawa A., Shibata K., Yoshino M., Itoh M., Ishii Y.,			
RA	Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S.,			
RA	Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamana K.,			
RA	Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R.,			
RA	Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T.,			
RA	Fleischmann W., Gaasterland T., Gissi C., King B., Kochava H.,			
RA	Kuehl P., Lewis S., Matsuo Y., Nikaido I., Pesole G., Quackenbush J.,			
RA	Schmitt L.M., Staudt F., Suzuki R., Tomita M., Wagner L., Washio T.,			
RA	Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Bash G.,			
RA	Blake J., Botelli D., Bojunga N., Carninci P., de Bonaldo M.F.,			
RA	Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M.,			
RA	Gustincich S., Hill D., Hofmann M., Hume D.A., Kamuya M., Lee N.H.,			

RA Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombaerts P.,
 RA Norioka P., Ring B., Ringwald M., Rodriguez I., Sakamoto N.,
 RA Sasaki H., Sato K., Schenbach C., Seya T., Shidara Y., Storch K.-F.,
 RA Suzuki H., Toyooka K., Wang K.H., Wetzler C., Whittaker C., Williams L.,
 RA Watanabe-Borja A., Yoshida K., Hasegawa Y., Kawai H., Kohsaki S.,
 RA Hayashizaki Y.,
 RT "Functional annotation of a full-length mouse cDNA collection."
 RL Nature 409:685-690(2001).
 DR EMBL: AF131236; AAD3016.1; -
 DR EMBL: AK091133; BAB26078.1; -
 DR MGD: MGI:1349479; Chst4.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 DR Transferase.
 KW
 SQ SEQUENCE 388 AA; 44694 MW; D1E9D7796DF8574D CRC64;

Query Match 100.0%; Score 2042; DB 11; Length 388;
 Best Local Similarity 100.0%; Pred. No. 5.7e-183;
 Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLKKGRLLMFLGSOVIYVAFIHMSVHRHLSQRESRPVPVAVLSSWRSSSPVQOL 60
 DB 1 MMLKKGRLLMFLGSOVIYVAFIHMSVHRHLSQRESRPVPVAVLSSWRSSSPVQOL 60
 QY 61 FCGHDPVFLMEPAMHVMWTFSTAMKLMHVRDLRSVFLCDMSVPDAVNPGRKOS 120
 DB 61 FCGHDPVFLMEPAMHVMWTFSTAMKLMHVRDLRSVFLCDMSVPDAVNPGRKOS 120
 QY 121 SLFQWQSRLCSAPVCDPFAHEISSPKHCKLLCGQPFDMVEKCRSHGFVLEVR 180
 DB 121 SLFQWQSRLCSAPVCDPFAHEISSPKHCKLLCGQPFDMVEKCRSHGFVLEVR 180
 QY 181 LSLQALYPLTDPSLNHVHLVRDPAVRSREHTTIELMVDSHIVLGOHLETKEEQ 240
 DB 181 LSLQALYPLTDPSLNHVHLVRDPAVRSREHTTIELMVDSHIVLGOHLETKEEQ 240
 QY 241 PYRAMKTIICKSOVDIYKAIOTLPALQOQRTFLRYEDLVRAPIAQTTRLYKFGDLPLH 300
 DB 241 PYRAMKTIICKSOVDIYKAIOTLPALQOQRTFLRYEDLVRAPIAQTTRLYKFGDLPLH 300
 QY 301 LOTWVYNTVRGKMGQAHFTNARNALNYSQAWRSLPYEKYSQLODAGCEAMDILGYLO 360
 DB 301 LOTWVYNTVRGKMGQAHFTNARNALNYSQAWRSLPYEKYSQLODAGCEAMDILGYLO 360
 QY 361 VRSQOEGNLSLDLSSSHILGOVFRG 388
 DB 361 VRSQOEGNLSLDLSSSHILGOVFRG 388

RESULT 2

Q9RL11 PRELIMINARY; PRT; 388 AA.
 AC Q9RL11;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE L-selectin ligand sulfotransferase.
 GN CHS74.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99361934; PubMed-10435581;
 RA Hirooka N., Petryniak B., Nakayama J., Tsuboi S., Suzuki M., Yeh J.C.,
 RA Izawa D., Tanaka T., Miyasaka M., Lowe J.B., Fukuda M.,
 RT "A novel, high endothelial venule-specific sulfotransferase expresses
 6-sulfo sialyl Lewis(x), an L-selectin ligand displayed by CD34.";
 RL Immunology 11:79-89(1999).
 DR EMBL: AF109155; AAD45579.1; -
 DR MGD: MGI:1349479; Chst4.
 DR InterPro: IPR000863; Sulfotransferase.

DR Pfam: PF00685; Sulfotransfer; 1.
 KW Lectin; Selectin; Transferase.
 SQ SEQUENCE 388 AA; 44636 MW; 6D53771AFB6884AEE CRC64;

Query Match 99.6%; Score 2033; DB 11; Length 388;
 Best Local Similarity 99.5%; Pred. No. 4.3e-183;
 Matches 386; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLKKGRLLMFLGSOVIYVAFIHMSVHRHLSQRESRPVPVAVLSSWRSSSPVQOL 60
 DB 1 MMLKKGRLLMFLGSOVIYVAFIHMSVHRHLSQRESRPVPVAVLSSWRSSSPVQOL 60
 QY 61 FCGHDPVFLMEPAMHVMWTFSTAMKLMHVRDLRSVFLCDMSVPDAVNPGRKOS 120
 DB 61 FCGHDPVFLMEPAMHVMWTFSTAMKLMHVRDLRSVFLCDMSVPDAVNPGRKOS 120
 QY 121 SLFQWQSRLCSAPVCDPFAHEISSPKHCKLLCGQPFDMVEKCRSHGFVLEVR 180
 DB 121 SLFQWQSRLCSAPVCDPFAHEISSPKHCKLLCGQPFDMVEKCRSHGFVLEVR 180
 QY 181 LSLQALYPLTDPSLNHVHLVRDPAVRSREHTTIELMVDSHIVLGOHLETKEEQ 240
 DB 181 LSLQALYPLTDPSLNHVHLVRDPAVRSREHTTIELMVDSHIVLGOHLETKEEQ 240
 QY 241 PYRAMKTIICKSOVDIYKAIOTLPALQOQRTFLRYEDLVRAPIAQTTRLYKFGDLPLH 300
 DB 241 PYRAMKTIICKSOVDIYKAIOTLPALQOQRTFLRYEDLVRAPIAQTTRLYKFGDLPLH 300
 QY 301 LOTWVYNTVRGKMGQAHFTNARNALNYSQAWRSLPYEKYSQLODAGCEAMDILGYLO 360
 DB 301 LOTWVYNTVRGKMGQAHFTNARNALNYSQAWRSLPYEKYSQLODAGCEAMDILGYLO 360
 QY 361 VRSQOEGNLSLDLSSSHILGOVFRG 388
 DB 361 VRSQOEGNLSLDLSSSHILGOVFRG 388

RESULT 3

Q9Y5R3 PRELIMINARY; PRT; 386 AA.
 AC Q9Y5R3;
 DT 01-NOV-1999 (TREMblrel. 12, Created)
 DT 01-NOV-1999 (TREMblrel. 12, Last sequence update)
 DT 01-MAR-2003 (TREMblrel. 23, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase (L-selectin ligand
 sulfotransferase GST-3).
 GN GST3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99264336; PubMed-10330415;
 RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
 RA Huang C.C., Kannagi R., Rosen S.D., Hemmerich S.,
 RT "Sulfotransferases of two specificities function in the reconstitution
 of high endothelial cell ligands for L-selectin.";
 RL J. Cell Biol. 145:899-910(1999).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-99264336; PubMed-10330415;
 RA Bistrup A., Bhakta S., Lee J.K., Belov Y.Y., Gunn M.D., Zuo F.R.,
 RA Hemmerich S., Lee J.K., Bhakta S., Bistrup A., Ruddle N.R.,
 RA Rosen S.D.;
 RT "Chromosomal Localization and Genomic Organization for the
 Galactose/N-Acetylglucosamine/N-Acetylglucosamine 6-O-
 sulfotransferase Gene Family.";
 RL Glycobiology 0:0-0(2001).
 RN [3]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-2132592; PubMed-11439191;
 RA Yeh J.-C., Hirooka N., Petryniak B., Nakayama J., Ellies L.G.,

RA Rabuka D., Hindsgaul O., March J.D., Lowe J.B., Fukuda M.;
 RT "Novel sulfated lymphocyte homing receptors and their control by a
 RT core extension lateral, 3-N-acetylglucosaminyltransferase.";
 RL Cell 105:957-969(2001).
 DR EMBL: AF131235; AAD33015.1;
 DR EMBL: AF280088; AAG48246.1;
 DR EMBL: AF149783; AAK48417.1;
 DR InterPro: IPR001092; HLH_basic.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00665; Sulfotransferase.
 DR PROSITE: PS00036; HLH_1; 1.
 DR Lectin: Selectin; Transferrase.
 SW SEQUENCE 386 AA; 45133 MW; 0C3BB4022417143A CRC64;

Query Match 74.1%; Score 1513.5; DB 4; Length 386;
 Best Local Similarity 72.8%; Pred. No. 7/7e-138;
 Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;

QY 1 MMLKGRLLMFLGQYIVVALFTIMSVHR--HLSOREESRRPVHVLVLSMRSGSSFFVG 58
 1 MLPLKMKKLLFLVSQAILALFFHMYSHNISLSMKAQPER-MHVLVLSMRSGSSFFVG 59
 DB 59 QLEQHPDVFYLMEPANHWMTFTSTAMKLHAAVRDLRSVFLCDMSVFDAVMNPPGRK 118
 60 QLEQHPDVFYLMEPANHWMTFTSTAMKLHAAVRDLRSVFLCDMSVFDAVMNPPGRK 119
 QY 119 QSSLFQWEOSSALCSAPVCEFPFAHETSPPKCKLGGQOPDMVEKACRSHGVYLVKEV 178
 120 QSSLFQWEOSSALCSAPVCEFPFAHETSPPKCKLGGQOPDMVEKACRSHGVYLVKEV 179
 DB 179 RFLSLQALYPLDPSLNLHVHLVRDPAVFRSREHTTIELMVDSHIVLGQHLLETKEE 238
 180 RFLSLQALYPLDPSLNLHVHLVRDPAVFRSREHTTIELMVDSHIVLGQHLLETKEE 239
 QY 239 DQPYAAKTIICKSQVDYKAIQTLPEALQRYLFLEKEDVLRAPLAQTRIKRYVGDFL 298
 240 DQPYAAKTIICKSQVDYKAIQTLPEALQRYLFLEKEDVLRAPLAQTRIKRYVGDFL 299
 DB 299 PHLOTWYVNTYRGKMGQAHFHTNARNALNYSQAMRSLPEYKESQLODACEAMDLIGY 358
 300 PHLOTWYVNTYRGKMGQAHFHTNARNALNYSQAMRSLPEYKESQLODACEAMDLIGY 359
 QY 359 LQVRSQOQDGNLSDLLSSHIIGQV 384
 360 RHVRSQOQDGNLSDLLSSHIIGQV 385
 DB

RESULT 4

Q8NCG5 PRELIMINARY; PRT; 386 AA.

AC Q8NCG5; PRT; 386 AA.
 DT 01-OCT-2002 (Tremblrel. 22, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Hypothetical protein FLJ90265.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Isogai T., Ota T., Nishikawa T., Hayashi K., Otsuki T., Sugiyama T.,
 RA Suzuki Y., Nagai K., Sugano S., Ishii S., Kawai-Hio Y., Saito K.,
 RA Yamamoto J., Wakamatsu A., Nakamura Y., Kojima S., Nagahara K.,
 RA Mesubi Y., Ono T., Okano K., Yoshikawa Y., Aotsuka S., Sasaki N.,
 RA Hattori A., Okumura K., Iwayanagi T., Nidomiya K.;
 RT "NEDO human cDNA sequencing project."
 RL Submitted (MAR-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AK074746; BAC11177.1;
 DR Genew: HGNC:1972; CHST4.
 KW Hypothetical protein.
 SQ SEQUENCE 386 AA; 45160 MW; 861869348319E42A CRC64;

Query Match 74.0%; Score 1510.5; DB 4; Length 386;
 Best Local Similarity 72.5%; Pred. No. 1.5e-137;
 Matches 280; Conservative 50; Mismatches 53; Indels 3; Gaps 2;

QY 1 MMLKGRLLMFLGQYIVVALFTIMSVHR--HLSOREESRRPVHVLVLSMRSGSSFFVG 58
 1 MLPLKMKKLLFLVSQAILALFFHMYSHNISLSMKAQPER-MHVLVLSMRSGSSFFVG 59
 DB 59 QLEQHPDVFYLMEPANHWMTFTSTAMKLHAAVRDLRSVFLCDMSVFDAVMNPPGRK 118
 60 QLEQHPDVFYLMEPANHWMTFTSTAMKLHAAVRDLRSVFLCDMSVFDAVMNPPGRK 119
 QY 119 QSSLFQWEOSSALCSAPVCEFPFAHETSPPKCKLGGQOPDMVEKACRSHGVYLVKEV 178
 120 QSSLFQWEOSSALCSAPVCEFPFAHETSPPKCKLGGQOPDMVEKACRSHGVYLVKEV 179
 DB 179 RFLSLQALYPLDPSLNLHVHLVRDPAVFRSREHTTIELMVDSHIVLGQHLLETKEE 238
 180 RFLSLQALYPLDPSLNLHVHLVRDPAVFRSREHTTIELMVDSHIVLGQHLLETKEE 239
 QY 239 DQPYAAKTIICKSQVDYKAIQTLPEALQRYLFLEKEDVLRAPLAQTRIKRYVGDFL 298
 240 DQPYAAKTIICKSQVDYKAIQTLPEALQRYLFLEKEDVLRAPLAQTRIKRYVGDFL 299
 DB 299 PHLOTWYVNTYRGKMGQAHFHTNARNALNYSQAMRSLPEYKESQLODACEAMDLIGY 358
 300 PHLOTWYVNTYRGKMGQAHFHTNARNALNYSQAMRSLPEYKESQLODACEAMDLIGY 359
 QY 359 LQVRSQOQDGNLSDLLSSHIIGQV 384
 360 RHVRSQOQDGNLSDLLSSHIIGQV 385
 DB

RESULT 5

Q8IV46 PRELIMINARY; PRT; 370 AA.

AC Q8IV46; PRT; 370 AA.
 DT 01-MAR-2003 (Tremblrel. 23, Created)
 DT 01-MAR-2003 (Tremblrel. 23, Last sequence update)
 DT 01-MAR-2003 (Tremblrel. 23, Last annotation update)
 DE Similar to carbonyldehydrate (N-acetylglucosamine 6-O) sulfotransferase
 4.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_Taxid=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Tissue-Brain;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DBJ databases.
 DR EMBL: BC05282; AAH35282.1;
 KW Transferase.
 SQ SEQUENCE 370 AA; 43249 MW; EA8D76EA4E73C625 CRC64;

Query Match 72.0%; Score 1470.5; DB 4; Length 370;
 Best Local Similarity 73.6%; Pred. No. 1e-133;
 Matches 271; Conservative 45; Mismatches 49; Indels 3; Gaps 2;

QY 19 VALFLHMSVHR--HLSOREESRRPVHVLVLSMRSGSSFFVGQLEQHPDVFYLMEPANH 76
 3 ILALFFHMYSHNISLSMKAQPER-MHVLVLSMRSGSSFFVGQLEQHPDVFYLMEPANH 61
 DB 77 VMKFTSTAMKLHAAVRDLRSVFLCDMSVFDAVMNPPGRKQSSLFQWEOSSALCSAPV 136
 62 VMKFTSTAMKLHAAVRDLRSVFLCDMSVFDAVMNPPGRKQSSLFQWEOSSALCSAPV 121
 QY 137 CDFPFAHETSPPKCKLGGQOPDMVEKACRSHGVYLVKEVRFSLQALYPLDPSLNL 196
 122 CDFPFAHETSPPKCKLGGQOPDMVEKACRSHGVYLVKEVRFSLQALYPLDPSLNL 181
 DB 197 LRVHLVRDPAVFRSREHTTIELMVDSHIVLGQHLLETKEEDPYAAKTIICKSQVDY 256
 197 LRVHLVRDPAVFRSREHTTIELMVDSHIVLGQHLLETKEEDPYAAKTIICKSQVDY 256

DB 182 LHIYHLVADPRAVRSRRTKGDMLIDSRIVMGHOKIKKEDOPYYMWQVYCOSOLETY 241

OY 257 KAIOTLPEALQOORLFLRYEDLVAPLAQOTRLRYKFGVGLDFPHLOTWYVNTTRGKMG 316

DB 242 KTIGSLPRALQERLYLVREEDLAPRAVQTSRMVEFGVLEPLPHLOTWVNTTRGKMG 301

OY 317 HAFHTNANALNVSAQMRWSLPEYKVSQLODAGCAMDLGLYQVRSQOEGNLSIDL 376

DB 302 HAFHTNADALNVSAQMRWSLPEYKVSRLQKACGAMMLLGYRHVRSRQORNLIDL 361

OY 377 SSHILQV 384

DB 362 TWYVEQI 369

RESULT 6

900P4 PRELIMINARY: PRT: 395 AA.

OC 01-MAY-2000 (TREMBLrel. 13, Created)

DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE N-acetylglucosamine 6-O-sulfotransferase.

GN CHST5 OR I-GLCNAC-6-ST.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_TaxID=10090;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=C57BL/6; TISSUE=Intestine;

RK MEDLINE=99423499; PubMed=10491328;

RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.,

RT "Cloning and Characterization of a Mammalian N-Acetylglucosamine-6-

RT Sulfotransferase that is Highly Restricted to Intestinal Tissue.";

RL Biochem. Biophys. Res. Commun. 263:543-549(1999).

DR EMBL: AF176841; AAD56003.1; -

DR EMBL: AF176840; AAD56002.1; -

DR MGI: 1931825; Chst5.

DR InterPro: IPR000863; Sulfotransferase.

DR Pfam: PF00685; Sulfotransfer; 1.

KW Transferase.

SO SEQUENCE 395 AA; 44537 MW; 3FDF71E43ED383BE CRC64;

Query Match 48.2%; Score 984; DB 11; Length 395;

Best Local Similarity 54.9%; Pred. No. 1.3e-86;

Matches 207; Conservative 44; Mismatches 114; Indels 12; Gaps 6;

OY 8 RLMEFGSOVIYVALFIHMSVHRHLSORESRP-----VHVLVLSWSSGSSFGOLF 61

DB 2 RLPRP-SSTVMLSLMTVGTGLVFLVSQVSPSPAGLGERVHVLVLSWSSGSSFGOLF 60

OY 62 GQHPDVFYLMPEAMHWMTFTSSTAMKLMAYRDLRSVFLCDMSVFPAYNMPGRKOS 121

DB 61 SOHPDVFYLMPEAMHWMTFTSSTAMKLMAYRDLRSVFLCDMSVFPAYNMPGRKOS 119

OY 122 LFGWQSAALCSAPVCDFFPAHEISSPKHCKLLCGQPFDMVEKCRSHGVYLVKEVRL 181

DB 120 LFGWQSAALCSAPVCDFFPAHEISSPKHCKLLCGQPFDMVEKCRSHGVYLVKEVRL 179

OY 182 SLQALYPLTDPSSLNHLVHVPDRPRAVRSRRTTELMVDSHIVLGOHETKEEOP 241

DB 180 NLQVLYPLTDPSSLNHLVHVPDRPRAVRSRRTTELMVDSHIVLGOHETKEEOP 238

OY 242 YYAKRIKCSQVDIVK-AIQTLPEALQOORLFLRYEDLVAPLAQOTRLRYKFGVGLDFPH 300

DB 239 LRVVVEVCRSHVRIEALHMKPRPFLQDRYLRVREEDLADPLVIRELYATGTGLPQ 298

OY 301 LQTVYNTTRGKMG--QAHFTNANALNVSAQMRWSLPEYKVSQLODAGCAMDLGLY 358

DB 299 LQTVYNTTRGKMG--QAHFTNANALNVSAQMRWSLPEYKVSQLODAGCAMDLGLY 358

OY 359 LQVRSQOEGNLSIDL 375

DB 359 RSVHSELQORDLSIDL 375

RESULT 7

90GZX3 PRELIMINARY: PRT: 395 AA.

AC 09GZX3;

DT 01-MAR-2001 (TREMBLrel. 16, Created)

DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)

DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)

DE N-acetylglucosamine 6-O-sulfotransferase GST-4beta (Corneal

DE N-acetylglucosamine 6-O-sulfotransferase).

GN GST4BETA OR CHST6.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.

OX NCBI_TaxID=9606;

RN [1]

RP SEQUENCE FROM N.A.

RA Hemmerich S., Lee J.-K., Bhakta S., Bistrup A., Ruddle N.R.,

RA Rosen S.D.;

RT "Chromosomal Localization and Genomic Organization for the

RT Galactose/N-Acetylglucosamine/N-Acetylglucosamine 6-O-

RT Sulfotransferase Gene Family.";

RL Glycobiology 0:0-0(2001).

RN [2]

RP SEQUENCE FROM N.A.

RA MEDLINE=20472330; PubMed=11017086;

RA Akama T.O., Nishida K., Nakayama J., Watanabe H., Fujiwara T.,

RA Nakamura T., Dots A., Kawasaki S., Inoue T., Maeda N., Yamamoto S.,

RA Ozaki K., Kinoshita S., Shimomura Y., Tanigami A., Fukuda M.N.;

RT "Macular corneal dystrophy type I and type II are caused by distinct

RT mutations in a new sulphotransferase gene.";

RL Nat. Genet. 26:237-241(2000).

DR EMBL: AF280086; AAG48244.1; -

DR EMBL: AF218990; AAG26325.1; -

DR EMBL: AF219991; AAG26327.1; -

DR GeneW: HGNC:6938; CHST6.

DR InterPro: IPR000863; Sulfotransferase.

DR Pfam: PF00685; Sulfotransfer; 1.

KW Transferase.

SO SEQUENCE 395 AA; 44098 MW; 433CA60248A46F7 CRC64;

Query Match 47.8%; Score 977; DB 4; Length 395;

Best Local Similarity 53.2%; Pred. No. 6.3e-86;

Matches 207; Conservative 43; Mismatches 97; Indels 42; Gaps 7;

OY 2 MLTKGRILMEFGSOVIYVALFIHMSVHRHLSQRESRRP-----VHVLVLSW 50

DB 14 LLLAQTFLLFLVLS-----RPGSSFPAGGARVHVLVLSW 49

OY 51 RSGSSFFVGOLFQHPDVFYLMPEAMHWMTFTSSTAMKLMAYRDLRSVFLCDMSV 110

DB 50 RSGSSFFVGOLFQHPDVFYLMPEAMHWMTFTSSTAMKLMAYRDLRSVFLCDMSV 109

OY 111 YNPPGRKOSLFGWQSAALCSAPVCDFFPAHEISSPKHCKLLCGQPFDMVEKCRSH 170

DB 110 YNPPGRKOSLFGWQSAALCSAPVCDFFPAHEISSPKHCKLLCGQPFDMVEKCRSH 168

OY 171 GVVYLVKEVRFSLQALYPLTDPSSLNHLVHVPDRPRAVRSRRTTELMVDSHIVL 230

DB 169 SHVYLVKEVRFSLQALYPLTDPSSLNHLVHVPDRPRAVRSRRTTELMVDSHIVL 228

OY 231 HLETKEDOPYYAKRIKCSQVDIVKAIQTL-PEALQOORLFLRYEDLVAPLAQOTTR 288

DB 229 N-GTVVVEVCRSHVRIEALHMKPRPFLQDRYLRVREEDLADPLVIRELYATGTGLPQ 286

OY 289 LKVEVGLDFLPHLOTWVNTTRGKMG--QAHFTNANALNVSAQMRWSLPEYKVSQLO 346

DB 287 LKVEVGLDFLPHLOTWVNTTRGKMG--QAHFTNANALNVSAQMRWSLPEYKVSQLO 346

OY 347 DAGCEAMDGLGLYQVRSQOEGNLSIDL 375

DB 347 ELCAGALQILGYRPRYSDEQRNLAIDL 375

RESULT 8

09GZS9 PRELIMINARY; PRT; 411 AA.

AC 09GZS9; Q9UBV3;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE Intestinal glucosyl-6-sulfotransferase (intestinal N-acetylglucosamine-6-O-sulfotransferase) (N-acetylglucosamine 6-O-sulfotransferase).
 GN CHST5 OR I-GLCNAC-6-ST.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC MEDLINE-20472330; PubMed-11017086;
 RX Akama T.O., Nishida K., Nakayama J., Matanabe H., Ozaki K., Nakamura T., Dots A., Kawasaki S., Inoue Y., Maeda N., Yamamoto S., Fujikura T., Thonar E.J., Shimomura Y., Kinoshita S., Tanigami A., Fukuda M.N.;
 RT "Macular corneal dystrophy type I and type II are caused by distinct mutations in a new sulphotransferase gene.";
 RL Nat. Genet. 26:237-241(2000).
 RN [2]
 RP SEQUENCE OF 22-411 FROM N.A.
 RC TISSUE-Intestine;
 RX MEDLINE-99423499; PubMed-10491328;
 RA Lee J.-K., Bhakta S., Rosen S.D., Hemmerich S.;
 RT "Cloning and characterization of a Mammalian N-Acetylglucosamine-6-sulfotransferase that is highly restricted to intestinal tissue.";
 RL Biochem. Biophys. Res. Commun. 263:543-549(1999).
 DR EMBL: AF246718; AAC28023.1; -
 DR EMBL: AF219991; AAG26326.1; -
 DR EMBL: AF176839; AAD56001.1; -
 DR EMBL: AF176838; AAD56000.1; -
 DR Gene: HGNC:1973; CHST5.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransferase; 1.
 KW Transference.
 SQ SEQUENCE 411 AA; 46160 MW; 97642D54BE926E06 CRC64;

Query Match

Best Local Similarity 55.68; Score 958.5; DB 4; Length 411;
 Matches 193; Conservative 42; Mismatches 87; Indels 25; Gaps 6;

42 VHVLYDSNRSGSSFGGOLFQHPDVEYLMPEPAHVMWMTSTAMKLMHVAVDLLRSVF 101
 63 VHVLYDSNRSGSSFGGOLFQHPDVEYLMPEPAHVMWMTSTAMKLMHVAVDLLRSVF 122
 102 LCDMSVDAVMNGPGRKQSLFQWQSRALCSAPVCDFFPAHRISSPKCKLKGQOPFD 161
 123 LCDMSVDAVM-POSNNLSAFNMWATSRALCSPPACAPRGITISQDVCKITCTQPF 161
 162 MVEKACRSHGFVYLKVEFPLSLALPLTDPESLNAHVHVLVDPRAVRSRHTTTELM 221
 182 LAREACRSYSHVLLKVEFPLSLALPLTDPESLNAHVHVLVDPRAVRSRHTTTELM 241
 222 VDSHYLVG-----OHLETKEDQPYAKKTIICKSQVDYVKAQTL--PEALQQR 270
 242 RDNGIYLVGNGKVEADPHRLIRE-----VCRSHVRAEA-ATLKPPPLNGRY 290
 271 LFLRYEDLVRAPLAQTTLRYLKVGLDFPLHQLQTVYVNTFGKMGQ--NAFHTNANALN 328
 291 RLRYFEDLARREPLAEIRALYAFGLTLTQLLEAMININITGSGIGRIEFHSSNARN 350
 329 VSGAMWSPLEKVEVSQLODAGCAADLLGYLQVRSQOEGCNLDDL 375
 351 VSGAMWSPLEKVEVSQLODAGCAADLLGYLQVRSQOEGCNLDDL 397

RESULT 9

09EP78 PRELIMINARY; PRT; 484 AA.

AC 09EP78;
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-MAR-2003 (TREMBLrel. 23, Last annotation update)
 DE N-acetylglucosamine 6-O-sulfotransferase GST-5 (N-acetylglucosamine-6-O-sulfotransferase) (4600013M07R1X protein) (Carbohydrate
 GN CHST7 OR GST5 OR 2600013M07R1X.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sclurognathi; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6;
 RX PubMed-10956661;
 RA Bhakta S., Barts A., Bowman K.G., Kao W.M., Polak I., Lee J.-K., Cook B.N., Bruenl R.E., Rosen S.D., Bertozzi C.R., Hemmerich S.;
 RT "Sulfation of N-Acetylglucosamine by Chondroitin 6-Sulfotransferase 2 (GST-5).";
 RL J. Biol. Chem. 275:40226-40234(2000).
 RN [2]
 RP SEQUENCE FROM N.A.
 RA Uchimura K., Muramatsu T.;
 RT "Identification and molecular characterization of a cDNA encoding a novel N-acetylglucosamine-6-O-sulfotransferase.";
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C57BL/6J; TISSUE-Embryo;
 RX MEDLINE-21085660; PubMed-11217851;
 RA Kawai J., Shingawa A., Shibata K., Yoshino M., Itoh M., Ishii Y., Arakawa T., Hara A., Fukunishi Y., Konno H., Adachi J., Fukuda S., Aizawa K., Izawa M., Nishi K., Kiyosawa H., Kondo S., Yamanaka I., Saito T., Okazaki Y., Gojobori T., Bono H., Kasukawa T., Saito R., Kadota K., Matsuda H.A., Ashburner M., Batalov S., Casavant T., Fleischmann W., Gaasterland T., Glisic C., King B., Kochiwa H., Kuehl P., Lewis S., Matsuo Y., Nikaide I., Pesole G., Quackenbush J., Schirral L.M., Staubli F., Suzuki R., Tomita M., Wagner L., Washio T., Sakai K., Okido T., Furuno M., Aono H., Baldarelli R., Barab G., Blake J., Boffelli D., Bojunga N., Carninci P., de Bonaldo M.F., Brownstein M.J., Bult C., Fletcher C., Fujita M., Gariboldi M., Lyons P., Marchionni L., Mashima J., Mazzarelli J., Mombarts P., Nordone P., Ring B., Ringwald M., Rodriguez I., Sakamoto N., Sasaki H., Sato K., Schoenbach C., Seyg T., Shibata Y., Storch K.-F., Suzuki H., Toyooka K., Wang K.H., Weitz C., Whitaker C., Wilming L., Wyshaw-Boris A., Yoshida K., Hasegawa Y., Kawai J., Kohsaki S., Hayashizaki Y.;
 RT "Functional annotation of a full-length mouse cDNA collection.";
 RL Nature 409:685-690(2001).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Kidney;
 RA Strausberg R.;
 RL Submitted (DEC-2001) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF280089; AAC48247.1; -
 DR EMBL: AB040710; BAB13769.1; -
 DR EMBL: AK011202; BAB27465.1; -
 DR EMBL: BC019204; AAH19204.1; -
 DR MGD: MGI:1891767; Chst7.
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransferase; 1.
 KW Transference.
 SQ SEQUENCE 484 AA; 54766 MW; 9B195537D7AB7193 CRC64;

Query Match

Best Local Similarity 29.28; Score 595.5; DB 11; Length 484;
 38.48; Pred: No. 6,7e-49;

[illegible]

Query Match	Best Local Similarity	Matches	Score	DB	Length	Indels	Gaps
29.0%	37.7%	51	593	4	486	54	10
Pred. No. 1.2e-48							
Conservative							
Mismatches							
136							

Db 218 GLVETIACERSCPPAIALALEACECKRYPVVYIKDVRLLDGLVPLLRBQGLNKVYOLE 277

Qy 204 RDPRAVERSREHTTELMVDS-----HIVIGQH-----LETIKKE 238

Db 278 RDPRAVHNSRLTKSROGLREISIOVLRTRGRDGRFHRVLLLAHGVGAPGCGSRALPAPRA 337

Qy 239 DQPYV---AMKICKISOVDYIAKIOTLPALQORYFLFLEEDYVRAPLQOTPLVYKFGV 295

Db 338 D--FFLTGALFETICAMLRDLFLFAGCAPRWMLRRRIYLRLEEDIVKQPRQQLRRLLRFSG 395

Qy 296 DFLPLQITWYVNVTRGKGGK--GHAFHTNARNALNVNQAWRMSLPEYKVSQOLDACGEAMD 354

Db 396 RALAALDLFALMLMTGCAAVGADRPHLISARDAREAHVARERLSRQOVQVEACAPAMR 455

Qy 355 LIGYLQVRSQOE-----QGNLSIDL 374

Db 456 LLAY--PRSGEEDGAEQPRGETPLEM 480

Query Match	Best Local Similarity	Matches 136;	Conservative 50;	Mismatches 130;	Indels 41;	Gaps 6;
43	HLVLSRSSRSSFFGOLFEGHPDVEYLMERPMHMYMCTTSTSTAKLHNAVYDLRSYFL	102	101	100	99	98
101	HIYHATKRTGSFFGELFNQHPDVFYLYDPWHMLQALYPGDASLGLCALDKMLRSFLR	160	159	158	157	156
103	CDMSVFADYMMPG-PRKO-----SSLPQEOSRALCSAPVCDFFPAHE-----ISSP	148	147	146	145	144
161	CDFSLRLRIADPGDGERAPDSANTJTMLEFMRNRNTKVCSPPLCPAPRAADGVLEVD	220	219	218	217	216
149	KHCKLLCGQPPDMEYKAKRSHGEVYLVKEVRFSLQALYPLITDPSLHVVHLYVDPRA	208	207	206	205	204
221	KACESTCPVSLRALEACRKYPPVYVINDVRLDLGLVPLRLDPLGLNLKVQLFRDPRA	280	279	278	277	276
209	VFRSEHTITELMDSHTVL-----GQULETI-----KEDDQY	242	241	240	239	238
281	VHNSLRKRGQLRRESIGVLTRORGDFHRYLLAHGVADRGGARALPSAPRADDFLT	340	339	338	337	336
243	YAMKICISQDYIVAIOITLPEALQOYLTLEEDYLVAPLAPQOTRLKFEVGLDPLPHQ	302	301	300	299	298
341	SALEYICCAWLRDLFTGTGAPAWLRKRLRLRIEYLVQPOQOLRLRLRFSGLRTLAID	400	399	398	397	396
303	TMVYVNTVRGKMG- -OAHFTNARNALNYSOAMRMSLPEYKVSQLODACGEANDLIGY	358	357	356	355	354
401	AFAPMTGTGSAVGADRPFLHSARDAKREAVHMRRELSQEQVROQVETACAPARLLAY	457	456	455	454	453

RT biosynthesis of 6-sulfo steryl Lewis X: Molecular cloning, chromosomal mapping, and expression in various organs and tumor cells.";
 RL J. Biochem. 124:670-678(1998).
 DR EMBL; AB014679; BAA34265.2;
 KW Transferrase.
 SO SEQUENCE 483 AA; 52787 MW; E27797D44931BA18 CRC64;
 Query Match 28.2%; Score 576.5; DB 4; Length 483;
 Best Local Similarity 35.1%; Pred. No. 4,6e-47;
 Matches 131; Conservative 65; Mismatches 122; Indels 55; Gaps 9;
 QY 44 VLIVSMRSGSSFGVQLFGQHPDVFYLMPPAHVMMFTSSSTAMKILHMAVRDLRSVFLC 103
 DB 120 VYVFTTMRSGSSFFELRNQNPFFLEPPVHWOKLYPGDAVSLQGAARMLSLALYRC 179
 QY 104 DMSVFDAYNMPGRKOS----SLFQWESRALCSAPVCDFFPAHE-----ISSPKCKLL 154
 DB 180 DLSVFLYSPAGSGGRNLTLTGIFGAATKVKVCSPLC---PAYRREVGVLVDRCK-K 235
 QY 155 CGQGFPMVEKAKRSHGFVYLKEVRFSLQALYPLITDPSLHVVHLVDRPRAVRSRE 214
 DB 236 CPQRLARFEEDCRKRYRTLVKGVFVDAVLAFLRLDPAIDLKVIHLVDRPRAVSSR- 294
 QY 215 HTTIELMVDSHVIGQHLETKED-----OPY-----AM 245
 DB 295 -----ISRHGLIRESLQVNSRDPRAHMPLEAAGHKLGAKKEGCGPADYHALGAM 348
 QY 246 KIICKSQVDIVAIQTL--PEALQORYLFLRYEDLVRAPLAQTTRLYKFGVGLDFLPHQ 302
 DB 349 EVICNS---MAKTLQALQPPDMLQGHYLVRYEDLVGPVTLIRVYDFVGLVSPEME 405
 QY 303 TWYVNTYRGKGQAHFHTNARNALNVSAQMRWSLPYEKVSQLODAGSAMDLGLQYR 362
 DB 406 QPALNMTSSGSSSKPFVVSARNATQANAMTALTFOQIKOVEECYOPMAVLGYERYN 465
 QY 363 SQOEOGNLSLDLL 375
 DB 466 SPEEVKDLSTLL 478
 RESULT 15
 QYAC5 PRELIMINARY; PRT: 530 AA.
 AC QYAC5; OGGZNS; QY6P2;
 DT 01-NOV-1999 (TREMBlrel. 12, Created)
 DT 01-MAR-2003 (TREMBlrel. 23, Last sequence update)
 DT 01-MAR-2003 (TREMBlrel. 23, Last annotation update)
 DE Long form of N-acetylglucosamine-6-O-sulfotransferase (GlcNAc6ST) (Carbohydrate sulfotransferase 2).
 GN GN6ST OR CHST2.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE-Brain;
 RX MEDLINE=98391845; PubMed=9722682;
 RA Uchimura K., Muramatsu H., Kanane T., Ogawa H., Yamakawa T., Fan Q., Mitsuoka C., Kannagi R., Habuchi O., Yokoyama I., Yamamura K., Ozaki T., Nakagawa A., Kadomatsu K., Muramatsu T.,
 RA "Human N-acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of 6-sulfo steryl Lewis X: Molecular cloning, chromosomal mapping, and expression in various organs and tumor cells.";
 RT J. Biochem. 124:670-678(1998).
 RL [2]
 RP SEQUENCE FROM N.A.
 RC TISSUE-umbilical vein endothelial cells;
 RX MEDLINE=99168906; PubMed=10049591;
 RA Li X., Tedder T.F.;
 RT "CHST1 and CHST2 sulfotransferases expressed by human vascular endothelial cells: cDNA cloning, expression, and chromosomal localization.";

RL Genomics 55:345-347(1999).
 RN [3]
 RP SEQUENCE OF 48-530 FROM N.A.
 RC TISSUE-Placenta;
 RX MEDLINE=20498786; PubMed=11042394;
 RA Sakaguchi H., Kitagawa H., Sugahara K.;
 RT "Functional expression and genomic structure of human N-acetylglucosamine-6-O-sulfotransferase that transfers sulfate to b-N-acetylglucosamine at the nonreducing end of an N-acetylglucosamine sequence.";
 RL Biochim. Biophys. Acta 1523:269-276(2000).
 DR EMBL; AB014680; BAA34266.2;
 DR EMBL; AF083066; AAD20981.1;
 DR EMBL; AB021125; BAB16887.1;
 DR EMBL; AB021124; BAB16886.1;
 DR InterPro: IPR000863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransfer; 1.
 KW Transferrase.
 SO SEQUENCE 530 AA; 57857 MW; A82CA227B9D5651B CRC64;
 Query Match 28.2%; Score 576.5; DB 4; Length 530;
 Best Local Similarity 35.1%; Pred. No. 5,2e-47;
 Matches 131; Conservative 65; Mismatches 122; Indels 55; Gaps 9;
 QY 44 VLIVSMRSGSSFGVQLFGQHPDVFYLMPPAHVMMFTSSSTAMKILHMAVRDLRSVFLC 103
 DB 167 VYVFTTMRSGSSFFELRNQNPFFLEPPVHWOKLYPGDAVSLQGAARMLSLALYRC 226
 QY 104 DMSVFDAYNMPGRKOS----SLFQWESRALCSAPVCDFFPAHE-----ISSPKCKLL 154
 DB 227 DLSVFLYSPAGSGGRNLTLTGIFGAATKVKVCSPLC---PAYRREVGVLVDRCK-K 282
 QY 155 CGQGFPMVEKAKRSHGFVYLKEVRFSLQALYPLITDPSLHVVHLVDRPRAVRSRE 214
 DB 283 CPQRLARFEEDCRKRYRTLVKGVFVDAVLAFLRLDPAIDLKVIHLVDRPRAVSSR- 341
 QY 215 HTTIELMVDSHVIGQHLETKED-----OPY-----AM 245
 DB 342 -----ISRHGLIRESLQVNSRDPRAHMPLEAAGHKLGAKKEGCGPADYHALGAM 395
 QY 246 KIICKSQVDIVAIQTL--PEALQORYLFLRYEDLVRAPLAQTTRLYKFGVGLDFLPHQ 302
 DB 396 EVICNS---MAKTLQALQPPDMLQGHYLVRYEDLVGPVTLIRVYDFVGLVSPEME 452
 QY 303 TWYVNTYRGKGQAHFHTNARNALNVSAQMRWSLPYEKVSQLODAGSAMDLGLQYR 362
 DB 453 QPALNMTSSGSSSKPFVVSARNATQANAMTALTFOQIKOVEECYOPMAVLGYERYN 512
 QY 363 SQOEOGNLSLDLL 375
 DB 513 SPEEVKDLSTLL 525
 Search completed: August 9, 2003, 16:44:22
 Job time : 36.5891 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:41:19 ; Search time 16.5426 Seconds
(Without alignments) 992.382 Million cell updates/sec

Title: US-09-645-078-4
Perfect score: 2042
Sequence: 1 MLLKKGRLLMFLGSGQIVV.....NLSDLLSSSHILGQVREG 388

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310858 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA.*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/1/1aa/PCUOS.COMB.pep.*
6: /cgn2_6/ptodata/1/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	1513.5	74.1	US-09-045-284A-2	Sequence 2, Appl1
2	1513.5	74.1	US-09-190-911-1	Sequence 1, Appl1
3	1448.5	70.9	US-09-786-240-11	Sequence 11, Appl1
4	579.5	28.4	US-09-263-023-2	Sequence 2, Appl1
5	579.5	28.4	US-09-471-867-2	Sequence 2, Appl1
6	573.5	28.1	US-09-263-023-4	Sequence 4, Appl1
7	573.5	28.1	US-09-471-867-4	Sequence 4, Appl1
8	527.5	25.8	US-09-015-188-2	Sequence 2, Appl1
9	497	24.3	US-08-899-514-2	Sequence 2, Appl1
10	496.5	24.3	US-08-655-878-2	Sequence 2, Appl1
11	93	4.6	US-09-612-204B-24	Sequence 24, Appl1
12	91	4.5	US-08-887-534A-45	Sequence 45, Appl1
13	91	4.5	US-09-527-431-45	Sequence 45, Appl1
14	86	4.2	US-09-612-204B-2	Sequence 2, Appl1
15	85.5	4.2	US-09-489-847-277	Sequence 277, App
16	85.5	4.2	US-09-786-240-9	Sequence 9, Appl1
17	85.5	4.2	US-09-489-847-133	Sequence 133, App
18	85.5	4.2	US-09-292-768-64	Sequence 64, Appl1
19	83	4.1	US-09-609-816-10	Sequence 10, Appl1
20	83	4.1	US-09-609-816-5	Sequence 5, Appl1
21	82	4.0	US-09-150-133-11	Sequence 11, Appl1
22	82	4.0	US-09-150-141-11	Sequence 11, Appl1
23	82	4.0	US-09-374-483-11	Sequence 11, Appl1
24	82	4.0	US-09-374-824-11	Sequence 11, Appl1
25	82	4.0	US-09-374-492-11	Sequence 11, Appl1
26	81	4.0	US-09-811-469-6	Sequence 6, Appl1
27	81	4.0	US-09-107-532A-5427	Sequence 5427, Ap

28	80	3.9	225	4	US-08-832-488-2	Sequence 2, Appl1
29	79.5	3.9	302	4	US-09-252-991A-22091	Sequence 22091, A
30	79.5	3.9	350	2	US-08-852-481-2	Sequence 2, Appl1
31	79.5	3.9	724	4	US-09-328-352-7710	Sequence 7710, Ap
32	79	3.9	177	4	US-08-469-260A-53	Sequence 53, Appl1
33	79	3.9	177	4	US-08-488-446-53	Sequence 53, Appl1
34	79	3.9	177	4	US-08-467-344A-53	Sequence 53, Appl1
35	79	3.9	522	4	US-09-302-620B-97	Sequence 97, Appl1
36	79	3.9	535	4	US-09-134-001C-3338	Sequence 3338, Ap
37	78.5	3.8	470	2	US-08-959-011-1	Sequence 1, Appl1
38	78.5	3.8	507	2	US-08-845-161A-4	Sequence 4, Appl1
39	78.5	3.8	507	3	US-09-270-751-4	Sequence 4, Appl1
40	78.5	3.8	507	4	US-09-168-218B-5	Sequence 5, Appl1
41	78	3.8	436	2	US-08-576-626A-47	Sequence 47, Appl1
42	77.5	3.8	304	4	US-09-609-816-4	Sequence 4, Appl1
43	77.5	3.8	512	4	US-09-302-620B-103	Sequence 103, App
44	77	3.8	265	4	US-09-328-352-4347	Sequence 4347, Ap
45	77	3.8	607	4	US-09-252-991A-21640	Sequence 21640, A

ALIGNMENTS

RESULT 1	US-09-045-284A-2	Sequence 2, Application US/09045284A
Patent No. 6265192		
GENERAL INFORMATION:		
APPLICANT: Bistrup, Annette		
APPLICANT: Rosen, Steven D.		
APPLICANT: Hemmerich, Stefan		
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3		
FILE REFERENCE: 6510-107US1		
CURRENT APPLICATION NUMBER: US/09/045, 284A		
CURRENT FILING DATE: 1998-03-20		
NUMBER OF SEQ ID NOS: 9		
SOFTWARE: FastSeq for Windows Version 4.0		
SEQ ID NO 2		
LENGTH: 386		
TYPE: PRT		
ORGANISM: Homo sapiens		
US-09-045-284A-2		
Query Match	74.1%; Score 1513.5; DB 3; Length 386;	
Best Local Similarity	72.8%; Pred. No. 8.9e-164;	
Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;		
QY	1 MLLKKGRLLMFLGSGQIVV--HLSQRESRPPVHVLVLSMRSGSFGV 58	Sequence 2, Appl1
DB	1 MLLKKGRLLMFLGSGQIVV--HLSQRESRPPVHVLVLSMRSGSFGV 58	Sequence 2, Appl1
QY	59 QLFQHPVFLTPMRAHVMWTFSTWKLHMAVRODLRSVFLCDMSVFPAVYNNPGRK 118	Sequence 4, Appl1
DB	60 QLFQHPVFLTPMRAHVMWTFSTWKLHMAVRODLRSVFLCDMSVFPAVYNNPGRK 118	Sequence 4, Appl1
QY	119 QSSLFQHPVFLTPMRAHVMWTFSTWKLHMAVRODLRSVFLCDMSVFPAVYNNPGRK 178	Sequence 4, Appl1
DB	120 QSSLFQHPVFLTPMRAHVMWTFSTWKLHMAVRODLRSVFLCDMSVFPAVYNNPGRK 178	Sequence 4, Appl1
QY	179 RFLSQALYPLLTPLSLNLVHVLVDRPRAVRSRHTTIELMVDSHIVLQHLTKKEV 238	Sequence 4, Appl1
DB	180 RFLSQALYPLLTPLSLNLVHVLVDRPRAVRSRHTTIELMVDSHIVLQHLTKKEV 238	Sequence 4, Appl1
QY	239 DQPIYAKIIRKSGQDIYKALQTLPEALQXLYPLRYEDVLRAPLAQTRIKRYGDFL 298	Sequence 4, Appl1
DB	240 DQPIYAKIIRKSGQDIYKALQTLPEALQXLYPLRYEDVLRAPLAQTRIKRYGDFL 298	Sequence 4, Appl1
QY	299 PHLQTVYVNTVRGSGMGOHAFHTNARNLANSOAWRSPLPYEKYSOLODAGCEAMDLLGY 358	Sequence 4, Appl1
DB	300 PHLQTVYVNTVRGSGMGOHAFHTNARNLANSOAWRSPLPYEKYSOLODAGCEAMDLLGY 358	Sequence 4, Appl1
QY	359 LQVRSQDEQGNLSDLLSSSHILGQV 384	Sequence 4, Appl1
DB	359 LQVRSQDEQGNLSDLLSSSHILGQV 384	Sequence 4, Appl1

```

Db          360 RHVRSEQEQRLNLLDLSTWTVPEQI 385

RESULT 2
US-09-190-911-1
; Sequence 1, Application US/09190911
; Patent No. 635365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSTYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 1
; LENGTH: 386
; TYPE: PRT
; ORGANISM: H. sapiens
US-09-190-911-1

Query Match          74.1%; Score 1513.5; DB 4; Length 386;
Best Local Similarity 72.8%; Pred. No. 8.9e-164;
Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;

QY 1 MLLKKGRLLMFLGSOVIYVALFIHMSVHR--HLSQRESRRPVHYLVLSMRSGSSFG 58
DB 1 MLLPKMKLLPLVSOAMALLALFFHMYSHINISSLSKQAPER-MHVLVLSMRSGSSFG 59
QY 59 QLEFGHPDVFYLMEPAMHVMFTSTANKLHMAVDLRSVFLCDMSVFDAVMNPGPKR 118
DB 60 QLEFGHPDVFYLMEPAMHVMFTSTANKLHMAVDLRSVFLCDMSVFDAVMNPGPKR 119
QY 119 QSSLFQWMEOSRALCSAPVCDFFPAHEISSPKCKKLICGQPPDMVEKACRSHGVYLVKE 178
DB 120 QSSLFQWMEOSRALCSAPVCDFFPAHEISSPKCKKLICGQPPDMVEKACRSHGVYLVKE 179
QY 179 RLTSLOALYPLLTDPSSLNHHVHLVDRPRAVRSRREHTTIELMVDSHIVLQGHLETKEE 238
DB 180 RFLNQSILYPLTKDPSLNHHVHLVDRPRAVRSRREHTTIELMVDSHIVLQGHLETKEE 239
Y 239 DQPYAMKIICKSOVDIYKAIQTLPEALQORLYFLRYEDLVRAPLAQTTRLYKFGVGLDF 298
DB 240 DQPYAMKIICKSOVDIYKAIQTLPEALQORLYFLRYEDLVRAPLAQTTRLYKFGVGLDF 299
QY 299 PHLQTVWNVNTRGKGMGHAFTNARNALNVSQAMRMSLPYEKVSQLODAGEADMLIGY 358
DB 300 PHLQTVWNVNTRGKGMGHAFTNARNALNVSQAMRMSLPYEKVSQLODAGEADMLIGY 359
QY 359 LQVRSQEQGNLSLDLSSSHILGOV 384
DB 360 RHVRSEQEQRLNLLDLSTWTVPEQI 385

```

```

; TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
; FILE REFERENCE: PF-0592 PCI
; CURRENT APPLICATION NUMBER: US/09/786,240
; CURRENT FILING DATE: 2002-03-12
; PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,6
; PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PERL Program
; SEQ ID NO 11
; LENGTH: 386
; TYPE: PRT
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: Incyte ID No. 6558935 2617407CD1
US-09-786-240-11

Query Match          70.9%; Score 1448.5; DB 4; Length 386;
Best Local Similarity 71.1%; Pred. No. 2.3e-156;
Matches 275; Conservative 48; Mismatches 59; Indels 5; Gaps 4;

QY 1 MLLKKGRLLMFLGSOVIYVALFIHMSVHR--HLSQRESRRPVHYLVLSMRSGSSFG 58
DB 1 MLLPKMKLLPLVSOAMALLALFFHMYSHINISSLSKQAPER-MHVLVLSMRSGSSFG 59
QY 59 QLEFGHPDVFYLMEPAMHVMFTSTANKLHMAVDLRSVFLCDMSVFDAVMNPGPKR 118
DB 60 QLEFGHPDVFYLMEPAMHVMFTSTANKLHMAVDLRSVFLCDMSVFDAVMNPGPKR 119
QY 119 QSSLFQWMEOSRALCSAPVCDFFPAHEISSPKCKKLICG-QQFPDMVEKACRSHGVYLVKE 177
DB 120 QSSLFQWMEOSRALCSAPVCDFFPAHEISSPKCKKLICG-QQFPDMVEKACRSHGVYLVKE 178
QY 178 VRLTSLOALYPLLTDPSSLNHHVHLVDRPRAVRSRREHTTIELMVDSHIVLQGHLETKEE 237
DB 179 VRLTSLOALYPLLTDPSSLNHHVHLVDRPRAVRSRREHTTIELMVDSHIVLQGHLETKEE 238
QY 238 EDQPYAMKIICKSOVDIYKAIQTLPEALQORLYFLRYEDLVRAPLAQTTRLYKFGVGLDF 297
DB 239 EDQPYAMKIICKSOVDIYKAIQTLPEALQORLYFLRYEDLVRAPLAQTTRLYKFGVGLDF 298
QY 298 LPHLQTVWNVNTRGKGMGHAFTNARNALNVSQAMRMSLPYEKVSQLODAGEADMLIGY 357
DB 299 LPHLQTVWNVNTRGKGMGHAFTNARNALNVSQAMRMSLPYEKVSQLODAGEADMLIGY 358
QY 358 YLQVRSQEQGNLSLDLSSSHILGOV 384
DB 359 YHVRSEQEQRLNLLDLSTWTVPEQI 385

RESULT 4
US-09-263-023-2
; Sequence 2, Application US/09263023
; Patent No. 6037159
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadamatsu, Kenji
; APPLICANT: Kannagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/09/263,023
; CURRENT FILING DATE: 1999-03-05
; EARLIER APPLICATION NUMBER: JP 10-544007
; EARLIER FILING DATE: 1998-03-05
; EARLIER APPLICATION NUMBER: JP 10-177844
; EARLIER FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2

```


Db 296 ----TRSRHGLRESLQVVRSDPRHRMPLEAGHKLGAKEGVGPADYHALGMEV 351
 QY 248 ICKSOVDIVKAIQTL---PEALQORYFLRYEDLVRAPLAQTRLYKFEVGLDFLHLQTM 304
 Db 352 ICNS---WAKTLQTLALOPPDWLOGHYLVRYEDLVGDVYKTLRRRYDVGILLVSPEDMOF 408
 QY 305 VYNTVRGKMGQOHAHTNARNALNVSQAMRSLPEYKVSQLODACGEAMDLLGYLOVRSQ 364
 Db 409 ALNMTSGSGSSSKPFVVSARNAQTQANAMRFTALFQQLKQVEFEFYQPMAYLGYERVNSP 468
 QY 365 QEOGNLSIDL 375
 Db 469 EEVKDLSTLL 479

RESULT 7

-09-471-867-4
 Sequence 4, Application US/09471867
 Patent No. 6453289

GENERAL INFORMATION:
 APPLICANT: Uchimura, Kenji
 APPLICANT: Muramatsu, Hideki
 APPLICANT: Kadomatsu, Kenji
 APPLICANT: Kanagaki, Reiji
 APPLICANT: Habuchi, Osami
 APPLICANT: Muramatsu, Takashi
 TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-O-SULFOTRANSFERASE AND
 TITLE OF INVENTION: DNA ENCODING THE SAME
 FILE REFERENCE: TOYAMA1,001AUS
 CURRENT APPLICATION NUMBER: US/09/471,867
 EARLIER FILING DATE: 1999-12-23
 EARLIER APPLICATION NUMBER: US 09/263,023
 EARLIER FILING DATE: 1999-03-05
 EARLIER APPLICATION NUMBER: JP 10-54007
 EARLIER FILING DATE: 1998-03-05
 EARLIER APPLICATION NUMBER: JP 10-177844
 EARLIER FILING DATE: 1998-06-24
 NUMBER OF SEQ ID NOS: 10
 SOFTWARE: FASTSEQ for Windows Version 3.0
 SEQ ID NO 4
 LENGTH: 484
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-471-867-4

Query Match 28.1%; Score 573.5; DB 4; Length 484;
 Best Local Similarity 35.0%; Pred. No. 1.8e-56;
 Matches 130; Conservative 65; Mismatches 121; Indels 55; Gaps 9;

QY 46 VLSSRRSSSFVGLGFGHPDYFLMEPAHVMWMTFTSTAMKLMHAYRDLRSVFLCDM 105
 Db 123 VETTRSSSGSFGELEFNQNPVEFLYEPYHWMQILYRGDAVSLGGAARDMLSALYRDL 182
 QY 106 SFEDVYMPGRKOS----SLFQWESRALCSAPVCDPPAH-----ISSKHKLLCG 156
 Db 183 SVFQYXSPAGSGRLLTLTGFGATNKVYSSPLC---PAYRKEVGIYDDRCK-KCP 238
 QY 157 QQPFDMVEKACRSHGFVYLKEVRLSLQALYPLTDPSSLNHLVHVLVDRPRAVRSRHT 216
 Db 239 FQRLARFECECKRYTLVIKGVRFDAVLAFLDPLDLKVIHLVDRPRAVRSRHT 295
 QY 217 TIELMVDHIVLGOHLETKED-----QPY-----AMKI 247
 Db 296 ----TRSRHGLRESLQVVRSDPRHRMPLEAGHKLGAKEGVGPADYHALGMEV 351
 QY 248 ICKSOVDIVKAIQTL---PEALQORYFLRYEDLVRAPLAQTRLYKFEVGLDFLHLQTM 304
 Db 352 ICNS---WAKTLQTLALOPPDWLOGHYLVRYEDLVGDVYKTLRRRYDVGILLVSPEDMOF 408
 QY 305 VYNTVRGKMGQOHAHTNARNALNVSQAMRSLPEYKVSQLODACGEAMDLLGYLOVRSQ 364
 Db 409 ALNMTSGSGSSSKPFVVSARNAQTQANAMRFTALFQQLKQVEFEFYQPMAYLGYERVNSP 468

QY 365 QEOGNLSIDL 375
 Db 469 EEVKDLSTLL 479

RESULT 8

US-09-015-188-2
 Sequence 2, Application US/09015188C
 Patent No. 6393358

GENERAL INFORMATION:
 APPLICANT: Williams, Kevin J
 APPLICANT: Tabas, Ira
 TITLE OF INVENTION: A Human Gene Encoding Human Chondroitin
 TITLE OF INVENTION: 6-Sulfotransferase
 FILE REFERENCE: JEFF-0231
 CURRENT APPLICATION NUMBER: US/09/015,188C
 CURRENT FILING DATE: 1998-01-29
 NUMBER OF SEQ ID NOS: 17
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 2
 LENGTH: 411
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-015-188-2

Query Match 25.8%; Score 527.5; DB 4; Length 411;
 Best Local Similarity 35.6%; Pred. No. 2.4e-51;
 Matches 144; Conservative 63; Mismatches 155; Indels 43; Gaps 14;

QY 6 KGRLLMFGS---QYIVVALFIMSVH-----RHLSQR--EES-----RRVHYLV 46
 Db 6 KAVLLALALASIAIOYTAIRTFKSFHCPCGLAELRLCEESPTFAVNLSTRTHLI 65
 QY 47 LSSMRSSSFVGLGFGHPDYFLMEPAHVMWMTFTSTAMKLMHAYRDLRSVFLCDM 98
 Db 66 LATTSGSSSFVGLGFGHPDYFLMEPAHVMWMTFTSTAMKLMHAYRDLRSVFLCDM 125
 QY 99 SVFLCDMSVFDVYMPGRKOS--LFQWESRALCSAPVCDPPAH-----ISSKHKLLCG 155
 Db 126 SLYDCDLFLYENYIKPPVNHHTDRIFFRGASRVLCSPVCDPPADLVLEEGDCVRRK 185
 QY 156 GQPFDMVEKACRSHGFVYLKEVRLSLQALYPLTDPSSLNHLVHVLVDRPRAVRSRHT 215
 Db 186 GLNLTVAAEACRSHGFVYLKEVRLSLQALYPLTDPSSLNHLVHVLVDRPRAVRSRHT 245
 QY 216 TIELMVDHIVLGOHLETKEDQPY---YAMKICKSOVDIVKAIQTLPEALQORYL 271
 Db 246 T----FRDTYRLMRLMYGTGR---KPYNDVYQTLTVEDEDFNSYSTGLMRPMLKGYM 298
 QY 272 FLRYEDLVRAPLAQTRLYKFEVGLDFLHLQTMVYNTVRG-KMGQOHAHTNARNALNVS 330
 Db 299 LVRYEDLVRAPLAQTRLYKFEVGLDFLHLQTMVYNTVRG-KMGQOHAHTNARNALNVS 357
 QY 331 QAMRSLPEYKVSQLODACGEAMDLLGYLOVRSQEOGNLSIDL 375
 Db 358 EKMRFLSLDIYAFQNMCCQVLAQDGYKTIASSEELKPNVSLV 402

RESULT 9

US-08-899-514-2
 Sequence 2, Application US/08899514
 Patent No. 5910581

GENERAL INFORMATION:
 APPLICANT: FUKUDA, MASAKAZU
 APPLICANT: HABUCHI, OSAMI
 TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
 TITLE OF INVENTION: SULFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
 NUMBER OF SEQUENCES: 9
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: KNOBBE, MARTENS, OLSON & BEAR, LLP
 STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR

```

CITY: NEWPORT BEACH
STATE: CALIFORNIA
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899,514
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: DANIEL E ALTMAN
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: TOYAM21.001AUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714 760 0404
TELEFAX: 714 760 9502
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 479
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-899-514-2

```

Query Match	24.38;	Score 497;	DB 2;	Length 479;
-------------	--------	------------	-------	-------------

[illegible]

```

CORRESPONDENCE ADDRESS:
ADDRESSEE:
STREET:
CITY:
STATE:
COUNTRY:
ZIP:
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/655,878
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME:
REGISTRATION NUMBER:
REFERENCE/DOCKET NUMBER:
TELECOMMUNICATION INFORMATION:
TELEPHONE:
TELEFAX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 458
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-655-878-2

```

[illegible]

; TITLE OF INVENTION: Swine Gamma Herpesvirus DNA and Methods of Use
 ; FILE REFERENCE: 61/50-299
 ; CURRENT APPLICATION NUMBER: US/09/612,204B
 ; CURRENT FILING DATE: 2001-08-13
 ; PRIOR APPLICATION NUMBER: U.S. 60/142,736
 ; PRIOR FILING DATE: 1999-07-08
 ; PRIOR APPLICATION NUMBER: U.S. 60/168,532
 ; PRIOR FILING DATE: 1999-12-02
 ; NUMBER OF SEQ ID NOS: 36
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO: 24
 ; LENGTH: 865
 ; TYPE: prp
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Description of Artificial Sequence: Deduced amino
 ; OTHER INFORMATION: acid sequence of porcine gamma herpesvirus gpb
 ; OTHER INFORMATION: gene
 ; -09-612-204B-24

Query Match 4.6%; Score 93; DB 4; Length 865;
 Best Local Similarity 23.8%; Pred. No. 0.26;
 Matches 43; Conservative 20; Mismatches 50; Indels 68; Gaps 7;

QY 212 SREHTIELMDSHVIVGQ-----HLETIKEEDOPYAKTIICKSOVDIVKAI 259
 DB 571 SRPVYEFKNDSTIKYQQLGVNNEILLTLYLETCQENTERYF----- 614
 QY 260 QTLPEALQORYLFLRYEDLVRAPLAQTTRLYKFGVGLDFPLHQVWYVY-----TRG 311
 DB 615 ---QAKTMDYIKYKHEHKYPLSSITLLOFFILNF-----TLENDPFIETRYD 665
 QY 312 KGMGCHAFHTNARNALNVSQAMR-WSLPEKYS-----QLDAGCEAMDL 356
 DB 666 E-----KRLSNVEFDIETMREYNYAQRVSGLRKDLIDSTNRNPFVDAFGSLMDL 717
 QY 357 G 357
 DB 718 G 718

RESULT 12
 US-08-887-534A-45
 ; Sequence 45, Application US/08887534A
 ; Patent No. 6455323

GENERAL INFORMATION:

; APPLICANT: Holden, David W.
 ; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
 ; NUMBER OF SEQUENCES: 106
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 233 South Wacker Drive/6300 Sears Tower
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/887,534A
 ; FILING DATE:
 ; CLASSIFICATION: 536
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Rin-Laures, Li-Hsien
 ; REGISTRATION NUMBER: 33,547
 ; REFERENCE/DOCKET NUMBER: 28341/33996
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (312) 474-6300
 ; TELEFAX: (312) 474-0448
 ; TELE: (312) 474-6600

; INFORMATION FOR SEQ ID NO: 45:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 932 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-887-534A-45

Query Match 4.5%; Score 91; DB 4; Length 932;
 Best Local Similarity 20.0%; Pred. No. 0.49;
 Matches 70; Conservative 58; Mismatches 112; Indels 110; Gaps 18;

QY 23 FTHMSVHRLHLSQRESRRPVHYL-----VLSKRSQSSRYGQLGHPVFTIMEPA 74
 DB 621 FSHRAVHLDEQGTGYPLHHVPOKATFDIHNSPLSEAAVGEYGVNE---NKKK 676
 QY 75 WHVWMTFTSTAMKLMHMAVRDLRSVLCDS--VFDAVMNPGPKOSLFOEWSRALC 132
 DB 677 FNIWEAQIGDFA-----NMSQMTFDNFL-----FSSRSKMGERSGLT 713
 QY 133 SAPVCDFFP-AHEISSPKHCKL-----LCGOQFPDVEYKACRSHGFVYL-EVRFSL 183
 DB 714 L-----FLPHAYEGGQPEHSARLERFLQLAENNCTVNLSSSYFILRAQASLDS 768
 QY 184 QALYPLDPSLNLVHLYRPRVFRSREHTTIELMDSHIVLGQHETKEEDOPY 243
 DB 769 EQMRPL-----VMSPKSLRNK---TVAKRPID-BFTSGEPILTESYQADK 812
 QY 244 AKKIC---KSGVDIVKAIOTLPE-----ALQORYLFRYDVLRAPLAQTTRLYKFGV 294
 DB 813 YKVLILATGKMTIDLEKALAKNPDESIVLVALERLYPF--PBEETALIAQ----- 861
 QY 295 LDFLPHLQ--TWYVNTVRGKGQAHFHTNARNALNVSQAMRWSLPEKYV 342
 DB 862 ---LPHLEEVSWVQEEPKKQG-----AMLYVYPYV 890

RESULT 13
 US-09-527-431-45
 ; Sequence 45, Application US/09527431
 ; Patent No. 6485899

GENERAL INFORMATION:

; APPLICANT: Holden, David W.
 ; TITLE OF INVENTION: ANTI-BACTERIAL METHODS AND MATERIALS
 ; NUMBER OF SEQUENCES: 106
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray & Borun
 ; STREET: 233 South Wacker Drive/6300 Sears Tower
 ; CITY: Chicago
 ; STATE: Illinois
 ; COUNTRY: United States of America
 ; ZIP: 60606
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patent In Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/527,431
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/887,534
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Rin-Laures, Li-Hsien
 ; REGISTRATION NUMBER: 33,547
 ; REFERENCE/DOCKET NUMBER: 28341/33996
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (312) 474-6300
 ; TELEFAX: (312) 474-0448
 ; TELE: (312) 474-6600
 ; INFORMATION FOR SEQ ID NO: 45:

SEQUENCE CHARACTERISTICS:
 LENGTH: 932 amino acids
 TYPE: amino acid
 TOPOLOGY: linear
 MOLECULE TYPE: protein
 US-09-527-431-45

Query Match 4.5%; Score 91; DB 4; Length 932;
 Best Local Similarity 20.0%; Pred. No. 0.49;
 Matches 70; Conservative 58; Mismatches 112; Indels 110; Gaps 18;

QY 23 FTHMSVHRLSGRESRFPVHL-----VLSSNRSGSSFGQLFGHPVFLMEPA 74
 DB 621 FSHRAVHLDEQGTGYFLHHVDPQKATFDHNSPLSAAYVGFYGVNVE---NKS 676
 QY 75 WYVWMTFTSTWAKLHMVRLRLRSVFLDMS--VDAVMNNGPKRQSSLFQWEGSRALC 132
 DB 677 FNIWEAQYGDFA-----NMSQMTFDNFL-----FSSRSKRGERSGLT 713
 QY 133 SAPVCDFFP-AHEISSPKHCKL-----LCGQPFDMVEKACRSHGFVVLK-EVRFSL 183
 DB 714 L-----FLPHAYEGCGPENSARLERFLDLAENCTVYNLSSSYFHLIRQAASLDS 768
 Y 184 QALYPLDPSLNLAVHLVDRPRAVRSREHTTELMVDSHIVLGQHLETKEDQPY 243
 DB 769 EQMRPL-----VMSPKSLRNK---TVAKPID-EFTSGFEPILTESYQADK 812
 QY 244 AKKIC---KSGVDYKATQTLPE-----ALQGRFLFYREDLVAPLAQTRLYKFG 294
 DB 813 VTKVILATGKMTDLKALKNPDESULLVAIERLPF--PEEIEITALAQ----- 861
 QY 295 LDLPPLHQ--TWYVNTRGKMGQHAFTNARNALNVAQWRMSLPYERY 342
 DB 862 ---LPLLEVSVMQVEBKNG-----AKLVYRYVAV 890

RESULT 14
 US-09-612-204B-2
 ; Sequence 2, Application US/09612204B
 ; Patent No. 6461811

GENERAL INFORMATION:
 APPLICANT: Patience, Clive
 TITLE OF INVENTION: Swine Gamma Herpesvirus DNA and Methods of Use
 FILE REFERENCE: 61750-299
 CURRENT APPLICATION NUMBER: US/09/612,204B
 CURRENT FILING DATE: 2001-08-13
 PRIOR APPLICATION NUMBER: U.S. 60/142,736
 PRIOR FILING DATE: 1999-07-08
 PRIOR APPLICATION NUMBER: U.S. 60/168,532
 PRIOR FILING DATE: 1999-12-02
 NUMBER OF SEQ ID NOS: 36
 SOFTWARE: PatentIn Ver. 2.1
 SEQ ID NO 2
 LENGTH: 195
 TYPE: PRT
 ORGANISM: Artificial Sequence
 FEATURE:
 OTHER INFORMATION: Description of Artificial Sequence: Deduced amino
 OTHER INFORMATION: acid sequence derived from the first open reading
 OTHER INFORMATION: frame of the DNA of SEQ ID NO:1
 US-09-612-204B-2

Query Match 4.2%; Score 86; DB 4; Length 195;
 Best Local Similarity 26.5%; Pred. No. 0.16;
 Matches 26; Conservative 11; Mismatches 29; Indels 32; Gaps 2;

QY 212 SREHTTIELMVDSHIVLQ-----HLETKEDQPYAMKIKKSQVDYKAI 259
 DB 88 SRPVYTFKFMNDSTIKYKQGVNNEILLTYYLETCQENMEYF----- 131
 QY 260 QTLPEALQRYLFLKRELDVRAFLAQTIRLYKFGVGLDF 297
 DB 132 ---QAKTDMYIKNYEHLKTVPLSSITLDFIALNF 165

RESULT 15
 US-09-489-847-277
 ; Sequence 277, Application US/09489847
 ; Patent No. 6476195

GENERAL INFORMATION:
 APPLICANT: Rosen et al
 TITLE OF INVENTION: 98 Human Secreted Proteins
 FILE REFERENCE: P2031P1
 CURRENT APPLICATION NUMBER: US/09/489,847
 CURRENT FILING DATE: 2000-01-24
 EARLIER APPLICATION NUMBER: PCT/US99/17130
 EARLIER FILING DATE: 1999-07-29
 EARLIER APPLICATION NUMBER: 60/094,657
 EARLIER FILING DATE: 1998-07-30
 EARLIER APPLICATION NUMBER: 60/095,486
 EARLIER FILING DATE: 1998-08-05
 EARLIER APPLICATION NUMBER: 60/096,319
 EARLIER FILING DATE: 1998-08-12
 EARLIER APPLICATION NUMBER: 60/095,454
 EARLIER FILING DATE: 1998-08-06
 EARLIER APPLICATION NUMBER: 60/095,455
 EARLIER FILING DATE: 1998-08-06
 NUMBER OF SEQ ID NOS: 376
 SOFTWARE: PatentIn Ver. 2.0
 SEQ ID NO 277
 LENGTH: 380
 TYPE: PRT
 ORGANISM: Homo sapiens
 US-09-489-847-277

Query Match 4.2%; Score 85.5; DB 4; Length 380;
 Best Local Similarity 23.1%; Pred. No. 0.51;
 Matches 42; Conservative 24; Mismatches 77; Indels 39; Gaps 6;

QY 144 EISSPRCKLLCGQPFDMVEKACRSHGFVYLKEVRF-LSLQALYPLDPSLNLVHV- 201
 DB 198 KVKLKKTFLFLYRDFVRLISAFRS-----KFELENEEFYKRFVAPMLRLVANH 248
 QY 202 -LYRDRPRAVRSREHT-----IELMVDSHIVLGQHLETKEDQPYAMKIKKSQVDY 256
 DB 249 SLPASAREAFRAKLKVSFANFYIYLDLPH---EKLAPFNEHWRQYRLCHPCQIDYDFV 305
 QY 257 KATQTLPEALQRYLFLKRELDVRAFLAQTIRLYKFGVGLDF 296
 DB 306 GKLETTDEDAQQLQLQYDVRQLRPPPSYRNRKTAASSWEDWPAKIFLARQQLKYLEAD 365
 QY 297 FL 298
 DB 366 FV 367

Search completed: August 9, 2003, 16:45:38
 Job time: 18.5426 secs

THIS PAGE RI ANK (11/10/10)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Comphen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:42:44 ; Search time 20.553 seconds
(without alignments)
2241.953 Million cell updates/sec

Title: US-09-645-078-4

Perfect score: 2042
Sequence: 1 MMLTKGRLLMFLGSGVIVV.....NLSDLSSSHILGQYFRG 388

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 451899 seqs, 118759770 residues

Total number of hits satisfying chosen parameters: 451899

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published Applications_AA:*

- 1: /cgn2_6/ptodata/2/pubppa/US07_PUBCOMB.pep:*
- 2: /cgn2_6/ptodata/2/pubppa/PCT_NEW_PUB.pep:*
- 3: /cgn2_6/ptodata/2/pubppa/US06_NEW_PUB.pep:*
- 4: /cgn2_6/ptodata/2/pubppa/US06_PUBCOMB.pep:*
- 5: /cgn2_6/ptodata/2/pubppa/US07_NEW_PUB.pep:*
- 6: /cgn2_6/ptodata/2/pubppa/PCTUS_PUBCOMB.pep:*
- 7: /cgn2_6/ptodata/2/pubppa/US08_NEW_PUB.pep:*
- 8: /cgn2_6/ptodata/2/pubppa/US08_PUBCOMB.pep:*
- 9: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
- 10: /cgn2_6/ptodata/2/pubppa/US09_PUBCOMB.pep:*
- 11: /cgn2_6/ptodata/2/pubppa/US09C_PUBCOMB.pep:*
- 12: /cgn2_6/ptodata/2/pubppa/US09_NEW_PUB.pep:*
- 13: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
- 14: /cgn2_6/ptodata/2/pubppa/US10_PUBCOMB.pep:*
- 15: /cgn2_6/ptodata/2/pubppa/US10C_PUBCOMB.pep:*
- 16: /cgn2_6/ptodata/2/pubppa/US10_NEW_PUB.pep:*
- 17: /cgn2_6/ptodata/2/pubppa/US60_NEW_PUB.pep:*
- 18: /cgn2_6/ptodata/2/pubppa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1513.5	74.1	386	9	US-09-616-825-2
2	1513.5	74.1	386	14	US-10-007-262-1
3	1001	49.0	418	9	US-09-927-602-5
4	978	47.9	395	9	US-09-927-602-3
5	977	47.8	395	9	US-09-927-602-2
6	958.5	46.9	390	9	US-09-927-602-4
7	651.5	31.9	171	9	US-09-927-602-8
8	579.5	28.4	483	15	US-10-212-933-2
9	573.5	28.1	484	15	US-10-212-933-4
10	573.5	28.1	531	9	US-09-833-790-255
11	481.5	23.6	169	9	US-09-927-602-6
12	469.5	23.0	169	9	US-09-927-602-7
13	310	15.2	169	9	US-09-927-602-9
14	288.5	14.1	179	9	US-09-927-602-10
15	274	13.4	174	9	US-09-927-602-11

16	134.5	6.6	387	15	US-10-126-279-21	Sequence 21, Appl
17	93	4.6	865	14	US-10-055-364-24	Sequence 24, Appl
18	92	4.5	388	15	US-10-126-279-6	Sequence 6, Appl
19	92	4.5	500	10	US-09-731-872-466	Sequence 466, App
20	91.5	4.5	1169	9	US-09-815-242-13448	Sequence 13448, A
21	91	4.5	899	9	US-09-815-242-13346	Sequence 5356, Ap
22	91	4.5	932	9	US-09-815-242-12615	Sequence 12615, A
23	90.5	4.4	1169	9	US-09-815-242-13693	Sequence 13693, A
24	86	4.2	195	14	US-10-055-364-2	Sequence 2, Appl
25	85.5	4.2	414	10	US-09-978-295A-466	Sequence 466, App
26	85.5	4.2	414	10	US-09-764-857-12	Sequence 12, Appl
27	85.5	4.2	414	10	US-09-978-697-466	Sequence 466, App
28	85.5	4.2	414	10	US-09-978-192A-466	Sequence 466, App
29	85.5	4.2	414	10	US-09-999-832A-466	Sequence 466, App
30	85.5	4.2	414	11	US-09-978-189-466	Sequence 466, App
31	85.5	4.2	414	11	US-09-978-608A-466	Sequence 466, App
32	85.5	4.2	414	11	US-09-978-585A-466	Sequence 466, App
33	85.5	4.2	414	11	US-09-978-191A-466	Sequence 466, App
34	85.5	4.2	414	11	US-09-978-403A-466	Sequence 466, App
35	85.5	4.2	414	11	US-09-978-193A-466	Sequence 466, App
36	85.5	4.2	414	11	US-09-999-833A-466	Sequence 466, App
37	85.5	4.2	414	11	US-09-981-915A-466	Sequence 466, App
38	85.5	4.2	414	11	US-09-978-824-466	Sequence 466, App
39	85.5	4.2	414	11	US-09-918-585A-466	Sequence 466, App
40	85.5	4.2	414	11	US-09-978-423A-466	Sequence 466, App
41	85.5	4.2	414	11	US-09-978-193A-466	Sequence 466, App
42	85.5	4.2	414	11	US-09-999-830A-466	Sequence 466, App
43	85.5	4.2	414	11	US-09-978-757A-466	Sequence 466, App
44	85.5	4.2	414	11	US-09-978-187B-466	Sequence 466, App
45	85.5	4.2	414	11	US-09-978-643A-466	Sequence 466, App

ALIGNMENTS

RESULT 1
US-09-816-825-2
Sequence 2, Application US/09816825
Patent No. US20010051370A1
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107CON
CURRENT APPLICATION NUMBER: US/09/816, 825
CURRENT FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/045, 284
PRIOR FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 386
TYPE: PRT
ORGANISM: Homo sapiens
US-09-816-825-2

Query Match 74.1%; Score 1513.5; DB 9; Length 386;
Best Local Similarity 72.8%; Pred. No. 2.9e-151;
Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;
QY 1 MMLTKGRLLMFLGSGVIVV...NLSDLESRRPVHLVLSWRGSSFGV 58
1 MLLEPKKKLLFLVSQALILAFHMTSHNISSKMAQPER-NHVLVLSWRGSSFGV 59
QY QLEGFQHDVYVLEPAHVMVMTFTSPTAMKLMHVAVDLSVFLCDKSVFDAYNPPRK 118
60 QLEGFQHDVYVLEPAHVMVMTFTSPTAMKLMHVAVDLSVFLCDKSVFDAYNPPRK 119
QY 119 OSSLPQWESRALCSAPVDFPFAHETISPRKCKLLCGQDPFDMVERACSHGFTVKEY 178
120 OSSLPQWESRALCSAPVDFPFAHETISPRKCKLLCGQDPFDMVERACSHGFTVKEY 179

Oy	179	FELSLQALYPLITDPSLNLHVHVLKRPDPVRSRPHHTIETLMWOSHIYGOLHEIKKE	238
Dd	180	RFFNLSILXPLPKDDSLMLHHVLKRDPAVRSSRRRTGGDLMDISRIMWGEOQLKKE	239
Oy	239	DOPYAMKICKSQVDIVKAIOITLEPEALOQRRLFLREEDLVBPAPLAQTTRILXYEVLDELT	298
Dd	240	DOPYVMVGIOGSOLEIKTKTIGSLPKAQERLLRVREDLARPAPQTSRMVEFVDELFL	299
Oy	299	PHLOTWYNVNRGRKGMGCHAFHTNARNALINSQAMRWSLPYEKVSOLQDNCGEMLDLG	358
Dd	300	PHLOTWINIRTRGKMGMGHAFHTNARDALINSAWMRSLPYEKVSRILQACGDANMILGY	359
Oy	359	LQVRSQOEGLNSLIDTLSSSHILGOV	384
Dd	360	RHVRSSEQRNLLEDLSTWTVPDOI	385

```

      SEQUENCE: 1
      PUBLICATION: US10007262
      GENERAL INFORMATION:
      APPLICANT: Bistrup, Annette
      APPLICANT: Rosen, Steven D.
      APPLICANT: Tangemann, Kirsten
      APPLICANT: Hemmerich, Stefan
      TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
      FILE REFERENCE: 6510-107C1P
      CURRENT APPLICATION NUMBER: US/10/007,262
      CURRENT FILING DATE: 2001-11-08
      PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
      PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
      NUMBER OF SEQ ID NOS: 8
      SOFTWARE: FastSeq for Windows Version 3.0
      SEQ ID NO: 1
      LENGTH: 386
      TYPE: prt
      ORGANISM: H. sapiens
      US-10-007-262-1
    
```

Query Match	74.1%	Score 1513.5	DB 14	Length 386
Best Local Similarity	72.8%	Pred. No. 2.9e-151		
Matches 281	Conservative .49	Mismatches 33	Indels 3	Gaps 2
QY	1	MLLKGRLMLFLGSOVIVVALFTHMSVHR	-HLSQRESRRPVHVLVLSWSRSGSFPVG	58
DB	1	MLERKMKLLFLVSOAMILLFLFHMYSHNISLISMSKQPER	-MHVLLVSSWRSGSSFPVG	59
QY	59	QLFQOHPRVFILMPAMHVMNTFTFSSYAKLHMAVROLRLRSYFLCDMSVFADYAMPGRK		118
DB	60	QLFQOHPRVFILMPAMHVMNTFKOSTYMMMLHMAVRDILRAVFLCDMSVFADYAMBERGR		119
QY	119	QSLFLFOWESRALCSAPVCDFFPAHEISSPKHCKLQCOQFPDWEKACRSHGFWLKEV		178
DB	120	QSLFLFOWENSALCSAPACDIIIPODEIILPRAHCRLLCSQOQPREVEYERKACRSYHVLKEV		179
QY	179	RFLSLQALPLLTDBSLNLAVYHLVDRPRAVFRSEHTTIELANDSHVYLGQHETKEE		238
DB	180	RFNPLQSLYPLPKPSPMLNTHVHLVDRPRAVFRSEHRTKGLDMSRTYMGQHEQKKEE		239
QY	239	DQPYAMKILICKSOVDIYKALQTLPEALQOORLFLRYEDIVARAPLQOTRLTKFPFGDPL		298
DB	240	DQPYVMQVYICOSQLEIKTQISLPKALQERLLVRYEDILARAPAOISRKMEFVGLTEFL		299
QY	299	PHLOTWYVNTYRGKMGQHAFTNARNALNYSQAWRMSLPEKYSQOLDACGEANDLILGY		358
DB	300	PHLOTWYVNTYRGKMGDHAFTNARNALNYSQAWRMSLPEKYSRLQKACGDANLLGY		359
QY	359	LQVRSQOQGNLSDLLSSHHILGY	384	
DB	360	RHVRSSEQORNLDDLSTWTPVEPI	385	

```

: RESULT 3
: US-09-927-602-5
: Sequence 5, Application US/09927602
: Patent No. US20020061562A1
: GENERAL INFORMATION:
: APPLICANT: Fukuda, Michiko N.
: TITLE OF INVENTION: Methods of Treating Macular Corneal
: TITLE OF INVENTION: Dystrophy
: FILE REFERENCE: P-LJ 4852
: CURRENT APPLICATION NUMBER: US/09/927,602
: CURRENT FILING DATE: 2001-08-09
: PRIOR APPLICATION NUMBER: US 09/638,211
: PRIOR FILING DATE: 2000-08-11
: NUMBER OF SEQ ID NOS: 38
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO 5
: LENGTH: 418
: TYPE: prt
: ORGANISM: Mus musculus
: US-09-927-602-5

```

Query Match	49.0%	Score 1001	DB 9	Length 418
Best Local Similarity	54.5%	Pred. No. 4.6e-97		
Matches	207	Conservative	45	Mismatches 110; Indels 10; Gaps 4
QY	4	LKGRLLMFLGSOYIVVALFIHMSVHRLSQREESRRD-----VHVLLTSMWSGSSFFV	57	
DB	21	LARGMRLEPRFSSTVALLSLMVTOTGILVFLVSROQVSSPAGLGERVHVLLVLSMRGSSFFV	80	
QY	58	GLRGQRHNDVYVLMERPAHVMKMTFTSSPAKMLHMAVRLLLSVFLCDMSYVDAYANNGPR	117	
DB	81	GOLFSQHDVYVLMERPAHVMDTLSQSGAPRLHMAVRLLISVFLCDMDVDYATL-PVRR	139	
QY	118	KQSLFQEOGSRALCSAPVCDFFPAHEISPKHCKLGGQPFDMVEKACSHGFEVKE	177	
DB	140	NISDLFGMAVRALCSPPVCEAFARGNISSEBCKPLCATRPFGLAQCACSSYSHVLAKE	199	
QY	178	VRFSLQALVPLLDPSLNLNVHLVDRPRAVFSREHTTELMADSHIVLGOHLETFKE	237	
DB	200	VRFPLQVLYPLLDSPALNLRIVHLVDRPRAVLRSREOTAFARLADNDSIVLGTN-GTAVE	258	
QY	238	EDQPYPAKKIICKSVDIVKAIQTLPRALOORYLELREDI-VRAPLAQTRTLRYKPVGDFE	297	
DB	259	ADPRLRVNEVCKSRIVRLAEALHMKPRPLQDRIKLVRIEDLARDPVLYIRELIYAFGTGL	318	
QY	298	LPHLOTWVYNTYRGKMG--QHAFTNARNALNVSQAWMSLPYEKVSQOLDACGEAMD	355	
DB	319	TPQLQGTWYHNTHGSGPGARREAFKTTSRDALSVSQAWRHTLPEAKIRRVDELGGALQ	378	
QY	356	LGELYDVRSQOEGQNSLDLL	375	
DB	379	LGRRVHSELEQRDSDIDL	398	

```

RESULT 4
US-09-927-602-3
; Sequence 3 'Application US/09927602
; Patent No. US20020061562A1.
; GENERAL INFORMATION:
; APPLICANT: Akuma, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Cornea)
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-IJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; CURRENT FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 3
; LENGTH: 395

```

TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: synthetic construct
NAME/KEY: VARIANT
LOCATION: (1)...(395)
OTHER INFORMATION: Xaa = any amino acid
US-09-927-602-3

Query Match 47.9%: Score 978; DB 9; Length 395;
Best Local Similarity 53.2%: Pred. No. 1.1e-94;
Matches 207; Conservative 42; Mismatches 98; Indels 42; Gaps 7;

QY 2 MLKKGRLLMFLGSOVIVLALFIMSVHRHLSQREESRRP-----VHVLVLSW 50
DB 14 LLLAQFTLLFLFVS-----RQPSPPAGEKRVHVLVLSW 49
QY 51 RSGSSFVGOLFQGHDPVFLMEPAHVMWMTFTSTANKLHMAVRDLRSVFLCDMSVFA 110
DB 50 RSGSSFVGOLFQGHDPVFLMEPAHVMWMTFTSTANKLHMAVRDLRSVFLCDMSVFA 109
QY 111 YMNGPRKQSSLPQWESRALCSAPVCDFFPAHISSPKCKLKGQPFDMVKACRSH 170
DB 110 YL-PWRRLSDLFQMAVSRAICSPACSAFPGXISSEYCKPLCARQPFKLAREACRSY 168
QY 171 GFVVLKEVRFLSLQALYPLTDPSLNTHVHLVNDPRAVFRSREHTTIELVDSHIVLQ 230
DB 169 SHVVLKEVRFLNQLVPLSDPALNRIYHLVNDPRAVLRSGQTAKALARDNGIYLG 228
QY 231 HLETKEDQPYAMKTIKCSQVDIVAKIOTL--PEALQORYLEFLREYDLVRAPLAQTTR 288
DB 229 N-GTWVADPDLRVAVREYCSHVRIAE-ATLKPPFLRGYRLVREDLAREPLAEIRA 286
QY 289 LYKRVGDLFLPHLOTWYNTYTRGKMG--QHAFTNARNALNVSQAWMSLPYKVSQLO 346
DB 287 LYAFETGLTPQLAEAMVHNTHTGSGPGARREAFKTSRNALNVSQAWMHLPFAKIRRVQ 346
QY 347 DACGAMDILGYLQVRSQOEGNLSIDL 375
DB 347 ELCAQALQLGLGRVYSEDEQRNLALDLV 375

RESULT 5
US-09-927-602-2
Sequence 2, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akema, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 2
LENGTH: 395
TYPE: PRT
ORGANISM: Homo Sapien
US-09-927-602-2

Query Match 47.8%: Score 977; DB 9; Length 395;
Best Local Similarity 53.2%: Pred. No. 1.4e-94;
Matches 207; Conservative 43; Mismatches 97; Indels 42; Gaps 7;

QY 2 MLKKGRLLMFLGSOVIVLALFIMSVHRHLSQREESRRP-----VHVLVLSW 50
DB 14 LLLAQFTLLFLFVS-----RQPSPPAGEKRVHVLVLSW 49
QY 51 RSGSSFVGOLFQGHDPVFLMEPAHVMWMTFTSTANKLHMAVRDLRSVFLCDMSVFA 110

DB 50 RSGSSFVGOLFQGHDPVFLMEPAHVMWMTFTSTANKLHMAVRDLRSVFLCDMSVFA 109
QY 111 YMNGPRKQSSLPQWESRALCSAPVCDFFPAHISSPKCKLKGQPFDMVKACRSH 170
DB 110 YL-PWRRLSDLFQMAVSRAICSPACSAFPGXISSEYCKPLCARQPFKLAREACRSY 168
QY 171 GFVVLKEVRFLSLQALYPLTDPSLNTHVHLVNDPRAVFRSREHTTIELVDSHIVLQ 230
DB 169 SHVVLKEVRFLNQLVPLSDPALNRIYHLVNDPRAVLRSGQTAKALARDNGIYLG 228
QY 231 HLETKEDQPYAMKTIKCSQVDIVAKIOTL--PEALQORYLEFLREYDLVRAPLAQTTR 288
DB 229 N-GTWVADPDLRVAVREYCSHVRIAE-ATLKPPFLRGYRLVREDLAREPLAEIRA 286
QY 289 LYKRVGDLFLPHLOTWYNTYTRGKMG--QHAFTNARNALNVSQAWMSLPYKVSQLO 346
DB 287 LYAFETGLTPQLAEAMVHNTHTGSGPGARREAFKTSRNALNVSQAWMHLPFAKIRRVQ 346
QY 347 DACGAMDILGYLQVRSQOEGNLSIDL 375
DB 347 ELCAQALQLGLGRVYSEDEQRNLALDLV 375

RESULT 6
US-09-927-602-4
Sequence 4, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akema, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 4
LENGTH: 390
TYPE: PRT
ORGANISM: Homo Sapien
US-09-927-602-4

Query Match 46.9%: Score 958.5; DB 9; Length 390;
Best Local Similarity 55.6%: Pred. No. 1.3e-92;
Matches 193; Conservative 42; Mismatches 87; Indels 25; Gaps 6;

QY 42 VHVLVLSWRSRSGSFVOLFQGHDPVFLMEPAHVMWMTFTSTANKLHMAVRDLRSV 101
DB 42 VHVLVLSWRSRSGSFVOLFQGHDPVFLMEPAHVMWMTFTSTANKLHMAVRDLRSV 101
QY 102 LCDMSVFDAYMNPGRKQSSLPQWESRALCSAPVCDFFPAHISSPKCKLKGQPF 161
DB 102 LCDMSVFDAYMNPGRKQSSLPQWESRALCSAPVCDFFPAHISSPKCKLKGQPF 160
QY 162 MVEKACRSHGCVYLVKEVFLSLQALYPLTDPSLNTHVHLVNDPRAVFRSREHTTIELM 221
DB 161 LARACRSYSHVVLKEVRFLNQLVPLSDPALNRIYHLVNDPRAVLRSGQTAKALARDNGIYLG 220
QY 222 VDSHIVG-----QHAFTNARNALNVSQAWMSLPYKVSQLO 270
DB 221 RDNGIIVGTNCKWEADPHRLIRE-----VCRSHVRIAE-ATLKPPFLRGY 269
QY 271 LFLRYEDLVRAPLAQTTRLYKRVGDLFLPHLOTWYNTYTRGKMG--QHAFTNARNALN 328
DB 270 LFLRYEDLVRAPLAQTTRLYKRVGDLFLPHLOTWYNTYTRGKMG--QHAFTNARNALN 329
QY 329 VSOAWMSLPYKVSQLODACGAMDILGYLQVRSQOEGNLSIDL 375
DB 330 VSOAWMSLPYKVSQLODACGAMDILGYLQVRSQOEGNLSIDL 376

```

RESULT 7
US-09-927-602-8
; Sequence 8, Application US/09927602
; Patent No. US20020061562A1
; GENERAL INFORMATION:
; APPLICANT: Fukuda, Michiko N.
; APPLICANT: Akama, Tomoya O.
; TITLE OF INVENTION: Methods of Treating Macular Corneal
; TITLE OF INVENTION: Dystrophy
; FILE REFERENCE: P-LJ 4852
; CURRENT APPLICATION NUMBER: US/09/927,602
; PRIOR FILING DATE: 2001-08-09
; PRIOR APPLICATION NUMBER: US 09/638,211
; PRIOR FILING DATE: 2000-08-11
; NUMBER OF SEQ ID NOS: 38
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 171
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-8

```

```

Query Match 31.9%; Score 651.5; DB 9; Length 171;
Best Local Similarity 47.2%; Pred. No. 11e-60;
Matches 134; Conservative 21; Mismatches 14; Indels 115; Gaps 3;

QY 32 LSQRESRPVHVLSSMRSGSFVQGLFGQHPDVFYLMPEPAMHVMFTSTAMKLM 91
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 3 LSKMQRPER-MHVLTLSSMRSGSFVQGLFGQHPDVFYLMPEPAMHVMFT----- 51

QY 92 AVRDLIRSVFLCDMSVFDAYNMPPRKQSSLFQWEOSRALCSAPVCDFFPAHEISSPKHC 151
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 52 ----- 51

QY 152 KLICGQPFDMVEKACRSHGFVYLKEVRFLSLQALYPLTDBSLNHVHLVDRPRAVR 211
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 52 -----KKACRSYSHVYLKEVRFLSLQALYPLTDBSLNHVHLVDRPRAVR 99

QY 212 SREHTTIELMVDSHVILGOHLETKEDQPYAMKICKSOYDIKAIOTLPEALQORYL 271
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 100 SRETKGKGLMDS-----KTISLKLALQERYL 127

QY 272 FLRYEDLVRAPIAOTRLKPYGLDFPLHLQVWYVNWTRKGMG 315
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 128 LVRYEDLARAAPVAYOTSRYEFGLEFLPHLQVWYVNWTRKGMG 171

```

```

RESULT 8
US-10-212-933-2
; Sequence 2, Application US/10212933
; Publication No. US20030008366A1
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kadamagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGALCOSAMINE-6-0-
; TITLE OF INVENTION: SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/10/212,933
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US/09/263,023
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: JP 10-54007
; PRIOR FILING DATE: 1998-03-05
; PRIOR APPLICATION NUMBER: JP 10-177844
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10

```

```

; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 483
; TYPE: PRT
; ORGANISM: Mus musculus
US-10-212-933-2

```

```

Query Match 28.4%; Score 579.5; DB 15; Length 483;
Best Local Similarity 35.4%; Pred. No. 1.9e-52;
Matches 132; Conservative 64; Mismatches 122; Indels 55; Gaps 9;

QY 44 VVLSMRSGSFVQGLFGQHPDVFYLMPEPAMHVMFTSTAMKLMHVAIRLLSVFLC 103
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 120 VVFTTMRSGSFVQGLFGQHPDVFYLMPEPAMHVMFTSTAMKLMHVAIRLLSVFLC 179

QY 104 DMSVFDAYNMPGPRKOS-----SLFQWEOSRALCSAPVCDFFPAHE-----ISSPKCKLL 154
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 180 DLSVFDAYSPPAGSGGRNLTTLGIFGATKRVKVCSSPLC---PAYKREYVGLVDRCK-K 235

QY 155 CGQQPPDMVEKACRSHGFVYLKEVRFLSLQALYPLTDBSLNHVHLVDRPRAVR 214
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 236 CPQRLAREEEDCRKRYRVIVGVDFVAVLAPLKDPLDLKYIHLVDRPRAVASSR- 294

QY 215 HTTIELMVDSHVILGOHLETKED-----OPRY-----FM 245
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 295 -----INSRHGLIRSLQVRSRDPRAHRMFFLEAGKLCAGKCGPADYHNLGAM 348

QY 246 KICKSOYDIKAIOTL---PEALQORYLFLRYEDLVRAPIAOTRLKPYGLDFPLHLQ 302
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 349 EVICNS---MAKTLQALQPPDMVGLQHYLVRYEDLVGPVATLRVYDFVGLVSPEME 405

QY 303 TWYVNTYTRKGMGQHNFTNARNALNVSOAMWSLPEYVSOLODAGCAGMILLGYOVR 362
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 406 QPALNMTSGSGSSKRFVVSARNATQANAMWRTALTFOQIKQVEECYQPMALGYERYN 465

QY 363 SOEOGNLSLDLL 375
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 466 SPEEVKDLSTKLL 478

```

```

RESULT 9
US-10-212-933-4
; Sequence 4, Application US/10212933
; Publication No. US20030008366A1
; GENERAL INFORMATION:
; APPLICANT: Uchimura, Kenji
; APPLICANT: Muramatsu, Hideki
; APPLICANT: Kadomatsu, Kenji
; APPLICANT: Kadamagi, Reiji
; APPLICANT: Habuchi, Osami
; APPLICANT: Muramatsu, Takashi
; TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGALCOSAMINE-6-0-
; TITLE OF INVENTION: SULFOTRANSFERASE AND
; FILE REFERENCE: TOYAMA1.001AUS
; CURRENT APPLICATION NUMBER: US/10/212,933
; PRIOR FILING DATE: 2002-08-05
; PRIOR APPLICATION NUMBER: US/09/263,023
; PRIOR FILING DATE: 1999-03-05
; PRIOR APPLICATION NUMBER: JP 10-54007
; PRIOR FILING DATE: 1998-03-05
; PRIOR APPLICATION NUMBER: JP 10-177844
; PRIOR FILING DATE: 1998-06-24
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 4
; LENGTH: 484
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-212-933-4

```

```

Query Match 28.1%; Score 573.5; DB 15; Length 484;
Best Local Similarity 35.0%; Pred. No. 8.4e-52;

```

Page 5

Query Match	28.1%	Score 573.5	DB 9	Length 531
Best Local Similarity	35.0%	Pred. No. 9.5e-52		
Matches 130	Conservative 65	Mismatches 121	Indels 55	Gaps 9
QY	46	VLSSMRSSSVGOLFQGHDPVFLYLMERAMHYMTFTSTAMKLMAYARDLRSFLCDM	105	
DB	170	VFTTRSSSSFEFGELFNQNEVEFLFLYELVWMMQKLVGODAVSLQGAARDMDLSALYKCDL	229	
QY	106	SVEFDAYMNPGRKOS---SLFQWESQALSCAPYCDFFPAHE-----ISSPKHCKLLCG	156	
DB	230	SVFOLYSPAGSGGRRLLTLLGIFGATNKNVVCSSPLC---PAYRKEVGVGLYDVRVCK-KCP	285	
QY	157	QQPEPMVEKACRSHGFVVLKEVRFSLQATVPLDPSLNLVHVLVDRPRAVFRSREHT	216	
DB	286	PQRLAREEBCRKTYTLVIKGVRFDAVLAFLDPLADLKVHILVNDPRAVYASR---	342	
QY	217	TIELMVDSHLYVGHLETKERD-----QPTV-----AMKI	247	

```

Query Match          23.6%; Score 481.5; DB 9; Length 169;
Best Local Similarity 38.3%; Pred. No. 9,7e-43;
Matches    105; Conservative   17; Mismatches   39; Indels   113; Gaps     3

OY      42 VHVVLVSSWBSGSSFVGQLEFGHPDYFLMEPAHWMMTFTTSSTAMKLMHVRDLRSVF 101
        |||||
DB       9 VHVVLVSSWSRSGSFFYGQLFNQHPPDVEFYLMBEPAHMWVT-----         47

OY      102 LCOMSVFDAQMNNGPRPKOSSLFQMWBESRALCAPVDCCFPFAHEISSPKNCKLLCGQPFD 161
        48 -----                                         47

OY      162 MYEKASRGHFVYLKEVERLSLOALYPLITDSNLNHHVVLRDPRAFRSRRETTIELM 221
           :||||| :||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
DB      48 -LSBACSYSHVYLKEVRFNFNQVLYRLPSDALMLRIHYHLRVDRBAVJRSREGTAALA 106

OY      222 VDHSIYLVGHLEPIKEEDOPYYIAMKTICKSQVDIVAKAIQTLPREALOORYLFLREDLVRA 281
           |:-|-|                                     |:-|-|
DB      107 RDNEAA-----TLKE-----PPFLGRGYRLVRFEEDARE 1359

OY      282 PLAGOTRYLKEVGLDFLPHLOTVMYVNTERGKGMG 315
           |||: ||| ||| :-|:-| ||| |||
DB      136 PLAETRILYAFTGLSLTPQEANRHNIHTHGSPGG 169

RESULT 12
US-09-927-602-7
Sequence 7, Application US/09927602
Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPlicant: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
             TITLE OF INVENTION: Dystrophy
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
```


Fri Aug 15 08:09:42 2003

us-09-645-078-4.rapb

Page 7

```

; SEQ ID NO 11
; LENGTH: 174
; TYPE: PRT
; ORGANISM: Homo Sapien
US-09-927-602-11

```

Query Match	13.48	Score 274	DB 9	Length 174
Best Local Similarity	24.78	Pred No	8.5e-21	
Matches 71	Conservative 29	Mismatches 57	Indels 130	Gaps 4

```

0Y 36 EESRPV-----HYLVSSMRSGSSFFVOLGQHBDVYLLMEPAHVMNTFSPAMKL 89
Db 1 EEPBPAAVAGPRRHVLMAITTTGSSFFVEEFNOCGNITLYLEPPLHIERTYS----- 53
0Y 90 HMAVBDLRSVFLCDMSVFDAYMNGPRKOSLSLFWQESRALCSAFVCDFFPAHEISSPK 149
Db 54 ----- 53
0Y 150 HCKLLCGQOPFDMVEKACSHGEVYVLKEVRFSLALAYPLTDPSLNLHVHLVRBPAY 209
Db 54 -----FEACRRKREHMLAKAVRINOLELOPLAEDPRDLKRYIQLOVRBPAY 100
0Y 210 FRSR-----EHTTELMDVSHIVLGHLTEKEBDOPYAMKIIICKSQVDIVKAIQTLP 263
Db 101 LASRWAFAGKYYKTWEJLGRQ-----P 122
0Y 264 EALQORLYLEFREDYLRAPLAOTRLYKRFVGLDPLHLOTWYNYNVR 310
Db 123 AMLKRRYMLVREEDVARGPLKAREMYFPAGLPLTFQVDEMDWIOKNTQ 169

```

Search completed: August 9, 2003, 16:46:25
Job time : 22.553 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:38:24 ; Search time 39.6021 Seconds
(without alignments)
1555.119 million cell updates/sec

Title: US-09-645-078-4
Perfect score: 2042
Sequence: 1 MMLKKGRLLMFISQVIVY.....NLSDLLSSHIILGVFREG 388

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1107863 seqs, 158726573 residues
Total number of hits satisfying chosen parameters: 1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
2: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
3: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
4: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
5: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
6: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
7: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
8: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
9: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
10: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1989.DAT.*
11: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*
24: /SIDSL/gcgdata/geneseq/geneseq-emb1/AA2003.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2042	100.0	388	20	AAV39919
2	1513.5	74.1	386	20	AAV39918
3	1510.5	74.0	386	22	AAV39918
4	1498.5	73.4	380	23	AAU11274
5	1468.5	70.9	386	21	AAV79219
6	1062	52.0	418	21	AAV81947
7	1001	49.0	418	21	AAV81557
8	984	48.2	395	22	AAV72638
9	984	48.2	395	23	AAV11275

10	978	47.9	395	23	ABB81555
11	977	47.8	395	22	AAV72640
12	977	47.8	395	23	ABB81554
13	977	47.8	395	23	AAE15438
14	958.5	46.9	390	22	AAV72639
15	958.5	46.9	390	22	ABB81556
16	651.5	31.9	171	23	ABB81560
17	592	29.0	486	24	ABP56121
18	579.5	28.4	483	20	AAV31656
19	576.5	28.2	530	22	ABB83567
20	573.5	28.1	484	20	AAV31657
21	573.5	28.1	531	19	AAV69414
22	527.5	25.8	411	19	AAV61100
23	527.5	25.8	411	23	AAE25356
24	527.5	25.8	411	24	ABU03503
25	507	24.8	499	24	ABR41139
26	497	24.3	479	19	AAV52863
27	496.5	24.3	458	18	AAV06480
28	481.5	23.6	169	23	ABB81558
29	469.5	23.0	169	23	ABB81559
30	310	15.2	169	23	ABB81561
31	288.5	14.1	179	23	ABB81562
32	274	13.4	174	23	ABB81563
33	210.5	10.3	363	22	ABB64512
34	176	8.6	183	22	ABB6582
35	134.5	6.6	388	22	AAE32782
36	130	6.4	1207	22	AAV72643
37	130	6.4	1207	24	AAE33542
38	126	6.2	596	22	AAV72641
39	126	6.2	1212	24	AAE33541
40	126	6.2	1222	22	AAV72642
41	126	6.2	1222	24	ABU11849
42	111	5.4	315	22	ABB64513
43	102	5.0	673	22	ABG18753
44	94.5	4.6	524	22	ABB59149
45	93	4.6	346	23	ABB91056

ALIGNMENTS

RESULT 1
AAV39919
ID AAV39919 standard; Protein: 388 AA.
AC AAV39919;
XX
XX
DT 08-DEC-1999 (first entry)
DE
DE
XX
KW Glycosyl sulfoltransferase; GST-3; detection; diagnosis; leukocyte homing;
KW selectin binding interaction; inflammation; lymphocyte homing; mouse;
KW secondary lymph organ.
XX
XX Mus. sp.
XX
XX W09949018-A1.
XX
XX
XX 30-SEP-1999.
XX
XX 26-FEB-1999; 99WO-US04316.
XX
XX 20-MAR-1998; 98US-0045284.
XX
XX 12-NOV-1998; 98US-0190911.
XX
XX (REGC) UNITEX CALIFORNIA.
XX (SYNT) SYNTX USA INC.
XX
XX Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
XX WPI: 1999-580442/49.
XX
XX N-PSDB; AA220793.

Consensus N-acetyl
Human glycosyl sul
Human corneal N-ac
Human drug metabo
Human glycosyl sul
Human intestinal N
Human high endothe
Human chondroitin
Mouse N-acetylgluc
Human protein sequ
Human N-acetylgluc
Lung small cell ca
Keratan sulphate 6
Human chondroitin
Angiogenesis-assoc
Human DITP enzyme
Glycosaminoglycan
Chick chondroitin
Human corneal N-ac
Human intestinal N
Human N-acetylgluc
Human keratan sul
Human chondroitin-
Drosophila melanog
Drosophila melanog
Mycobacterium tube
Mouse glycosyl sul
Human novel Cpe-as
Human novel Cpe-as
Human novel Cpe-as
Human glycosyl sul
Human secreted pro
Drosophila melanog
Novel human diagn
Drosophila melanog
Herbicide activ

XX Human and murine glycosyl sulfotransferase 3 and related
 PT polynucleotides
 XX
 PS Claim 2; Fig 4; 59pp; English.

XX This sequence is the mouse glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.

XX Sequence 388 AA;

Query Match 100.0%; Score 2042; DB 20; Length 388;
 Best Local Similarity 100.0%; Pred. No. 1.2e-208;
 Matches 388; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLKKGRLMFLGSGVIVVAFIMHSVHRHLSORESRRPVHVLVLSMRSGSPVGL 60
 DB 1 MMLKKGRLMFLGSGVIVVAFIMHSVHRHLSORESRRPVHVLVLSMRSGSPVGL 60
 QY 61 FGOHPDVFYLMPEPAHVMWMTFTSTAWKLMHMAVRDLRSVFLCDMSVFDAYMNPGRKOS 120
 DB 61 FGOHPDVFYLMPEPAHVMWMTFTSTAWKLMHMAVRDLRSVFLCDMSVFDAYMNPGRKOS 120
 QY 121 SLFQWEOBRALCSAPVCDFFPAHEISSPKHCKLGGQOPDMVEKACRSHGFVYLKEYE 180
 DB 121 SLFQWEOBRALCSAPVCDFFPAHEISSPKHCKLGGQOPDMVEKACRSHGFVYLKEYE 180
 QY 181 LSIQALYPLTDPISLHVVHLVDRPRAVRSREHTTIELMADSHVILGQHEETIKEEDQ 240
 DB 181 LSIQALYPLTDPISLHVVHLVDRPRAVRSREHTTIELMADSHVILGQHEETIKEEDQ 240
 QY 241 PYYAKTIICKSQVDIVKAIQTLPEALQORLFLREEDVLRAPLAQTRILKYFVGIDFLPH 300
 DB 241 PYYAKTIICKSQVDIVKAIQTLPEALQORLFLREEDVLRAPLAQTRILKYFVGIDFLPH 300
 QY 301 LOTWVYNTVRGKGMOHAFHTNARNALNVSOAMRWSLPYEKVSQLODACEAMDLGLGYLQ 360
 DB 301 LOTWVYNTVRGKGMOHAFHTNARNALNVSOAMRWSLPYEKVSQLODACEAMDLGLGYLQ 360
 QY 361 VRSQOEGNLSLDLSSSHILGQVFRFG 388
 DB 361 VRSQOEGNLSLDLSSSHILGQVFRFG 388

RESULT 2
 AAY39918
 ID AAY39918 standard; Protein; 386 AA.

AC AAY39918;
 DT 08-DEC-1999 (first entry)

DE Human glycosyl sulfotransferase-3 protein sequence.

XX glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
 KM selectin binding interaction; inflammation; lymphocyte homing; human;
 KW secondary lymph organ.

XX Homo sapiens.

XX WO9949018-A1.

XX 30-SEP-1999.

XX 26-FEB-1999; 99WO-US04316.

XX 20-MAR-1998; 98US-0045284.

PR 12-NOV-1998; 98US-0190911.
 XX
 PA (REGC) UNIV CALIFORNIA.
 PA (SYNT) SYNTX USA INC.

XX Bistrup A, Rosen SD, Tangemann K, Hemmerlich S;
 DR WPI; 1999-580442/49.
 XX N-PSDB; AA220792.

PT Human and murine glycosyl sulfotransferase 3 and related
 PT polynucleotides

PS Claim 2; Fig 1; 59pp; English.

CC This sequence is the human glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.

SQ Sequence 386 AA;

Query Match 74.1%; Score 1513.5; DB 20; Length 386;
 Best Local Similarity 72.8%; Pred. No. 2.6e-152;
 Matches 281; Conservative 49; Mismatches 53; Indels 3; Gaps 2;

QY 1 MMLKKGRLMFLGSGVIVVAFIMHSVHR--HLSORESRRPVHVLVLSMRSGSPVGL 58
 DB 1 MMLKKGRLMFLGSGVIVVAFIMHSVHR--HLSORESRRPVHVLVLSMRSGSPVGL 58
 QY 59 OLFQHPDVFYLMPEPAHVMWMTFTSTAWKLMHMAVRDLRSVFLCDMSVFDAYMNPGRK 118
 DB 59 OLFQHPDVFYLMPEPAHVMWMTFTSTAWKLMHMAVRDLRSVFLCDMSVFDAYMNPGRK 118
 QY 60 QLFQHPDVFYLMPEPAHVMWMTFTSTAWKLMHMAVRDLRSVFLCDMSVFDAYMNPGRK 119
 DB 60 QLFQHPDVFYLMPEPAHVMWMTFTSTAWKLMHMAVRDLRSVFLCDMSVFDAYMNPGRK 119
 QY 119 QSLFQWEOBRALCSAPVCDFFPAHEISSPKHCKLGGQOPDMVEKACRSHGFVYLKEYE 178
 DB 119 QSLFQWEOBRALCSAPVCDFFPAHEISSPKHCKLGGQOPDMVEKACRSHGFVYLKEYE 178
 QY 179 RELSIALYPLTDPISLHVVHLVDRPRAVRSREHTTIELMADSHVILGQHEETIKEEDQ 238
 DB 179 RELSIALYPLTDPISLHVVHLVDRPRAVRSREHTTIELMADSHVILGQHEETIKEEDQ 238
 QY 239 DQPYAMKTIICKSQVDIVKAIQTLPEALQORLFLREEDVLRAPLAQTRILKYFVGIDFL 298
 DB 239 DQPYAMKTIICKSQVDIVKAIQTLPEALQORLFLREEDVLRAPLAQTRILKYFVGIDFL 298
 QY 240 DQPYAMKTIICKSQVDIVKAIQTLPEALQORLFLREEDVLRAPLAQTRILKYFVGIDFL 299
 DB 240 DQPYAMKTIICKSQVDIVKAIQTLPEALQORLFLREEDVLRAPLAQTRILKYFVGIDFL 299
 QY 299 PHLQWVYNTVRGKGMOHAFHTNARNALNVSOAMRWSLPYEKVSQLODACEAMDLGLGY 358
 DB 299 PHLQWVYNTVRGKGMOHAFHTNARNALNVSOAMRWSLPYEKVSQLODACEAMDLGLGY 358
 QY 300 PHLQWVYNTVRGKGMOHAFHTNARNALNVSOAMRWSLPYEKVSQLODACEAMDLGLGY 359
 DB 300 PHLQWVYNTVRGKGMOHAFHTNARNALNVSOAMRWSLPYEKVSQLODACEAMDLGLGY 359
 QY 359 LQVRSQOEGNLSLDLSSSHILGQV 384
 DB 359 LQVRSQOEGNLSLDLSSSHILGQV 384
 QY 360 RHVRSEQEGNLSLDLSSSHILGQV 385
 DB 360 RHVRSEQEGNLSLDLSSSHILGQV 385

RESULT 3
 AAM93309
 ID AAM93309 standard; Protein; 386 AA.

AC AAM93309;
 DT 06-NOV-2001 (first entry)

DE Human polypeptide; SEQ ID NO: 2817.

XX Human; full length cDNA; cDNA synthesis; oligo-capping.

XX Homo sapiens.

PN EP1130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI: 2001-524255/58.
DR N-PSDB: AAK94229.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
XX Claim 8; SEQ ID NO 2817; 1380bp + sequence listing; English.
XX
XX The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been
XX isolated and nucleotide sequences of 5' - and 3' -ends of the cDNA
XX molecules have been determined. Primers for synthesizing the full length
XX cDNA are useful for clarifying the function of the protein encoded by
XX the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
XX method. The primers enable the production of the full length cDNA easily
XX without any special methods. The present sequence is a polypeptide
XX Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 386 AA;
Query Match 74.0%; Score 1510.5; DB 22; Length 386;
Best Local Similarity 72.5%; Pred. No. 5.4e-152;
Matches 280; Conservative 50; Mismatches 53; Indels 3; Gaps 2;
QY 1 MMLIKGRLLMFLGQYIVVALFTIMSVHR--HLSOREESRRPVHVLVLSMRSGSSFFVG 58
D 1 MMLPKKKLLFLVSOAAILALFFHMYSHNITSLSMKAQPER-MHVLVLSMRSGSSFFVG 59
QY 59 QLRGQHDVFTYLMRPAHVMWTFSTSTAMKILMAVRLDLSVFLCDMSVFDAYNPGPRK 118
D 60 QLRGQHDVFTYLMRPAHVMWTFSTSTAMKILMAVRLDLSVFLCDMSVFDAYNPGPRK 119
QY 119 QSLFQWESRALCSAVCDFEPFAHETSPKHKCLKGGOOPFDMVEKACRSHGVVLEKEY 178
D 120 QSLFQWESRALCSAVCDFEPFAHETSPKHKCLKGGOOPFDMVEKACRSHGVVLEKEY 179
QY 179 RFLSLQALYPLLDPSINLHVHLVNDPRAVFRSREHTTELMVDSHIVLGOHLETKEE 238
D 180 RFLSLQALYPLLDPSINLHVHLVNDPRAVFRSREHTTELMVDSHIVLGOHLETKEE 239
QY 239 DQPYAKTIICKSQVDVYKAIQITPEALQRYLEFREDEVRAVLAQTTRIKYVGDDEL 298
D 240 DQPYAKTIICKSQVDVYKAIQITPEALQRYLEFREDEVRAVLAQTTRIKYVGDDEL 299
QY 299 PHLTQVYNTYTRGKMGQAHFHTNARNALNVSQAMRSLSPEKYSQODACGEAMDLGY 358
D 300 PHLTQVYNTYTRGKMGQAHFHTNARNALNVSQAMRSLSPEKYSQODACGEAMDLGY 359
QY 359 LQYRSQOEGNLSLDLSSHIIGOV 384
D 360 RHVRSQOEGNLSLDLSSHIIGOV 385

RESULT 4
AAU11274

ID AU11274 standard; Protein; 380 AA.
XX
AC AU11274;
XX
DT 12-MAR-2002 (first entry)
XX
DE Human L-selectin sulfotransferase-2 (LSST-2) protein.
XX
KW Human; beta1,3GNT; beta1,3-N-acetylglucosaminyltransferase; MECA-79;
KW L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
KW ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
KW allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
KW delayed-type hypersensitivity reaction; hyperplastic thymus; antilucer;
KW antinflammatory; antipsoriatic; antidiabetic; dermatological;
KW antiallergic.
XX
OS Homo sapiens.
XX
PN WO200185177-A1.
XX
PD 15-NOV-2001.
XX
PF 10-MAY-2001; 2001WO-US15452.
XX
PR 11-MAY-2000; 2000US-0569320.
XX
PA (BURN-) BURNHAM INST.
PI Fukuda M, Yeh J, Hirooka N;
XX
DR WPI: 2002-075226/10.
DR N-PSDB: AAS16947.
XX
PT New enzyme, useful for modifying acceptor molecule, comprises an
PT isolated L-selectin sulfotransferase-2 that directs expression of
PT L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
PT intestinal G1CNAc 6-sulfotransferase -
XX
XX Claim 21; Fig 4; 98pp; English.
XX
XX The present invention provides a method of modifying an acceptor molecule
XX by contacting the acceptor with an isolated
XX beta1,3-N-acetylglucosaminyltransferase (beta1,3GNT) or an active
XX fragment, where beta1,3GNT directs expression of a MECA-79 antigen. The
XX invention also provides a method of treating or preventing an
XX L-selectin-mediated condition by reducing the expression or activity of a
XX beta1,3GNT that directs expression of a MECA-79 antigen. This can be done
XX by administering to the subject an oligosaccharide L-selectin antagonist
XX that inhibits binding of L-selectin to a MECA-79 antigen, for example by
XX administering antibody material that specifically binds beta1,3GNT,
XX and/or a beta1,3GNT antisense nucleic acid molecule. L-selectin
XX sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression.
XX Alternatively, the expression or activity of LSST-2 or its active
XX fragment can be reduced in combination with reducing the expression or
XX activity of beta1,3GNT. The method is useful for treating L-selectin
XX mediated conditions such as Crohn's disease and ulcerative colitis,
XX inflammatory disorders of the skin such as allergic contact dermatitis,
XX psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type
XX hypersensitivity reactions, diabetes and hyperplastic thymus. This
XX sequence represents human LSST-2.
XX
SQ Sequence 380 AA;
Query Match 73.4%; Score 1498.5; DB 23; Length 380;
Best Local Similarity 72.1%; Pred. No. 1e-150;
Matches 277; Conservative 48; Mismatches 51; Indels 3; Gaps 2;
QY 8 RLMLFLGQYIVVALFTIMSVHR--HLSOREESRRPVHVLVLSMRSGSSFFVQLFGQHP 65
D 2 RLMLFLGQYIVVALFTIMSVHR--HLSOREESRRPVHVLVLSMRSGSSFFVQLFGQHP 60
QY 66 DVFYLMRPAHVMWTFSTSTAMKILMAVRLDLSVFLCDMSVFDAYNPGPRKQSSSLFQW 125

Db 61 DVFYLMERAMHVMWTFKOSTAMMLMAVRDLIRAVFLCDMSVEDAYMEGPRRQSSFLQW 120
 QY 126 EOSRALCAPVCDFFPAHEISSPKCKILCGQPDFDWEKACRSHGFWYKKEVRLSLQA 185
 Db 121 ENSRALCAPACDIIIPDEIIPRACHRLCSQGFVEVEKACRSYSHVYLKKEVRFENLQS 180
 QY 186 LYPILTDPSLNLHVHLVRDPAVRSRREHTTIELMDSHYLGCHLETKEEDOPYAM 245
 Db 181 LYPILKDPSSLNLHVHLVRDPAVRSRREHTTIELMDSHYLGCHLETKEEDOPYAM 240
 QY 246 KIICKSOVDYKAIQTLPEALQOYTLFREDVLRAPLAQOTRLKYKFGLDPLPHLOQWV 305
 Db 241 QVICSQLEIKYKTIOSLPRKALQERYLVRVEDLAPAPVAQTSRMTEFVGLEFLPHLOQWV 300
 QY 306 YNVTGRKGMOGHAFHTNARNALNVSQAMRMSLPYERKVSQLODACEGEMDLGLQVRSQO 365
 Db 301 HNITRGKMGDAFHTNARNADALNVSQAMRMSLPYERKVSRLQACGDAMNLGVRHVRSPQ 360
 Db 366 EOGNLSLDLSSSHILGOV 384
 Db 361 EQRNLLDLSTWTVPEQI 379

RESULT 5

AAY79219
 ID AAY79219 standard; Protein; 386 AA.

AC AAY79219;

DT 19-JUN-2000 (first entry)

DE Human transferase TRNSFS-11.

XX Transferrase; TRNSFS-11; human; antitumor; cell proliferation;
 KW gastrointestinal disorder; developmental disorder;
 KW genetic disorder; neurological disorder; reproductive disorder;
 KW smooth muscle disorder; immunological disorder; inflammation;
 KW diagnosis; therapy; N-acetylglucosamine 6-O-sulfotransferase.

XX Homo sapiens;

XX Key Location/Qualifiers

FH Modified-site 121
 FT /note= "potential O-phosphorylation"
 FT 107
 FT Modified-site /note= "potential O-phosphorylation"
 FT 217
 FT Modified-site /note= "potential O-phosphorylation"
 FT 252
 FT Modified-site /note= "potential O-phosphorylation"
 FT 364
 FT Modified-site /note= "potential O-phosphorylation"
 FT 380
 FT Modified-site /note= "potential O-phosphorylation"
 FT 35
 FT Modified-site /note= "potential O-phosphorylation"
 FT 50
 FT Modified-site /note= "potential O-phosphorylation"
 FT 81
 FT Modified-site /note= "potential O-phosphorylation"
 FT 287
 FT Modified-site /note= "potential O-phosphorylation"
 FT 243
 FT Modified-site /note= "potential O-phosphorylation"
 FT 30
 FT Modified-site /note= "potential O-phosphorylation"
 FT 308
 FT Modified-site /note= "potential N-glycosylation"
 FT 329
 FT Modified-site /note= "potential N-glycosylation"
 FT 7..23
 FT Domain /note= "potential N-glycosylation"
 XX /note= "transmembrane domain"

PN W0200044251-A2.
 XX 16-MAR-2000.
 PD 09-SEP-1999; 99WO-US20989.
 XX 10-SEP-1998; 98US-0150657.
 PR 04-NOV-1998; 98US-0186779.
 PR 11-MAY-1999; 99US-0133642.
 XX (INCYTE PHARM INC.
 PA Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H,
 PI Hillman JL, Azimzal Y;
 DR WPI: 2000-256996/22.
 DR N-PSDB; AA294211.
 XX Human transferase proteins useful for preventing, diagnosing and
 PT treating cancers and developmental, gastrointestinal, genetic,
 PT immunological, neurological, reproductive and smooth muscle disorders -
 PS Claim 1; Page 90-91; 113pp; English.

CC The present sequence is that of human transferase TRNSFS-11, 1 of
 CC 15 claimed novel human transferase proteins of the invention (see
 CC AAY79209-23). The sequence was deduced from a cDNA clone (see
 CC AA294211) isolated from a gallbladder library. It shows homology to
 CC mouse N-acetylglucosamine 6-O-sulfotransferase. TRNSFS-11 is
 CC expressed in dermatologic and gastrointestinal tissues,
 CC especially those associated with inflammation and cell
 CC proliferation. The new human transferases and polynucleotides can
 CC be used in the diagnosis, prevention and treatment of cancer,
 CC developmental disorders, gastrointestinal disorders, genetic
 CC disorders, immunological disorders, neurological disorders,
 CC reproductive disorders, and smooth muscle disorders. The
 CC polypeptides can also be used to raise antibodies, and to screen
 CC for agonists and antagonists of transferase activity.

SQ Sequence 386 AA;

Query Match 70.9%; Score 1448.5; DB 21; Length 386;

Best Local Similarity 71.1%; Pred. No. 2.2e-145;

Matches 275; Conservative 48; Mismatches 59; Indels 5; Gaps 4;

QY 1 MMLKRGRLFLMGSOVIVALLFTHMSVHR--HLSQRESRPPVHVLSSRRSSFFVG 58
 Db 1 MLLPKRMKLLFLVSGMALLALFFHYSNHSLSMKAPQR-MHVLVSSWRSSFFVG 59
 QY 59 QLFQGHDPVFLMEPRAMHVMWTFKOSTAMMLMAVRDLIRAVFLCDMSVEDAYMEGPRR 118
 Db 60 QLFQGHDPVFLMEPRAMHVMWTFKOSTAMMLMAVRDLIRAVFLCDMSVEDAYMEGPRR 119
 QY 119 GSSLFQWESRALCAPVCDFFPAHEISSPKCKILCGQPDFDWEKACRSHGFWYKKE 177
 Db 120 GSSLFQWESRALCAPACDIIIPDEIIPRACHRLCSQGFVEVEKACRSYSHVYLKKE 178
 QY 178 VRELSQALYPLTDPSSLNLHVHLVRDPAVRSRREHTTIELMDSHYLGCHLETKE 237
 Db 179 VREFNLQSLYPLKDPSSLNLHVHLVRDPAVRSRREHTTIELMDSHYLGCHLETKE 238
 QY 238 EDQPYTAAKTIKSOVDYKAIQTLPEALQOYTLFREDVLRAPLAQOTRLKYKVGDF 297
 Db 239 EDQPYTAAKTIKSOVDYKAIQTLPEALQOYTLFREDVLRAPLAQOTRLKYKVGDF 298
 QY 298 LPHLOQWVYNNVTGRKGMOGHAFHTNARNALNVSQAMRMSLPYERKVSRLQACGDAMNLG 357
 Db 299 LPHLOQWVYNNVTGRKGMOGHAFHTNARNADALNVSQAMRMSLPYERKVSRLQACGDAMNLG 358
 QY 358 YLQVRSQOEGNLSLDLSSSHILGOV 384
 Db 359 YRHRVSEQORNLLDLSTWTVPEQI 385

RESULT 6
 AAB41947
 ID AAB41947 standard; Protein: 418 AA.
 XX
 AC AAB41947;
 XX
 DT 08-FEB-2001 (first entry)
 XX
 DE Human ORFX ORE1711 polypeptide sequence SEQ ID NO: 3422.
 XX
 KW Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
 KW vlnervary; antiparasitic; antiparkinsonian; neotropic; neuroprotective;
 KW anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antirheumatic; antihypertensive;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 XX
 OS Homo sapiens.
 XX
 FN MO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PE 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Ieach M;
 DR WPI: 2000-602362/57.
 DR N-PSDB; AAC76156.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 CS Claim 11; Page 2599-2600; 5507pp; English.
 XX
 AC AACT74446 to AACT77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vlnervary;
 CC antiparasitic; antiparkinsonian; neotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypertensive; antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX
 SQ sequence 418 AA;
 Query Match 52.0%; Score 1062; DB 21; Length 418;
 Best Local Similarity 59.9%; Pred. No. 3.9e-104;
 Matches 208; Conservative 41; Mismatches 74; Indels 24; Gaps 5;
 QY 42 VHYIVLSSWRSGSSFGVQLGQHPDVFYLMPEAMHVMYMTSSPAMKLHMAVDLRSVF 101
 DB 69 VHYIVLSSWRSGSSFGVQLGQHPDVFYLMPEAMHVMYMTSSPAMKLHMAVDLRSVF 128
 QY 102 LCDMSVDAVMNPGPRKSSLFQWEGSRALCSAPVCDFFPAHEISSPKHKLCSGQPF 161
 DB 129 LCDMDVDVDAIMEPEPRRQSSLFQWENRALSAPACILITODEIIPRARRLCSGQPF 188
 QY 162 MEKACRSHGFVVAKEVRFSLQALYPLLDPSLNLVHVLVDRPRAVFRSREHTTELM 221
 DB 189 VVERKACRSHGVVLAKEVRFNLSQSLYPLLDPSLNLHIVLVRDPRAVFRSREMAAPILA 248
 QY 222 VDSHIVG-----QHLETKEDQPYAMKICKSOVDYKATQTL--PEALQORY 270
 DB 249 RDNQIVLGTGKWEADPHRLINE-----VRSVRIEA-ATLKPPFLGKRY 297
 QY 271 LFRYEDLVAPLAOTTRLYKFGVGLDPLPLQTVVYVTRGKMGQ--HAFHTNARNALN 328
 DB 298 RLVRFEDLVAREPLAEIRALYAFGLTLPLQLEAMIHNTGSGIGKPIEAFHTSSRNARN 357
 QY 329 VSOAMRWSLFEYKVSQLODAGCEAMDILGTVLSQOEGNLSIDL 375
 DB 358 VSOAMRHALPFTKILRVQEVCAQALQLLGRPYVSADQORDLTDLV 404
 RESULT 7
 ABB81557
 ID ABB81557 standard; Protein: 418 AA.
 XX
 AC ABB81557;
 XX
 DT 05-SEP-2002 (first entry)
 XX
 DE Mouse intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:5.
 XX
 KW Human; N-acetylglucosamine-6-sulfotransferase; enzyme; GlcNAc6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological.
 XX
 OS Mus musculus.
 XX
 FN US2002061562-A1.
 XX
 PD 23-MAY-2002.
 XX
 PR 09-AUG-2001; 2001US-0927602.
 PR 11-AUG-2000; 2000US-325773P.
 PA (FUKU/) FUKUDA M. N.
 PA (AKAM/) AKAMA T. O.
 XX
 PI Fukuda MN, Akama TO;
 DR WPI: 2002-507643/54.
 XX
 PT New nucleic acid encoding corneal
 PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,
 PT monitoring and diagnosis of macular corneal dystrophy -
 XX
 CS Example 5; Page 24-25; 69pp; English.
 CC The present invention describes human corneal
 CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyse
 CC sulfation of keratan sulfate (KS). Also described is a method for
 CC monitoring the effect of treatments for macular corneal dystrophy (MCD),

CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
 CC and has ophthalmological activity. (I) can be used to treat or prevent
 CC macular dystrophy types I or II. (II) makes possible treatment
 CC of MCD without requiring keratinoplasty or keratectomy. The present
 CC sequence represents mouse intestinal N-acetylglucosamine-6-
 CC sulfotransferase, which is given in comparison with (I) in the
 CC exemplification of the present invention.

CC
 XX
 SQ Sequence 418 AA;

Query Match 49.0%; Score 1001; DB 23; Length 418;
 Best Local Similarity 54.5%; Pred. No. 1,2e-97;
 Matches 207; Conservative 45; Mismatches 118; Indels 10; Gaps 4;

QY 4 LKKGRLMFLGSOVYVALFIHMSVHRHLSORESRP-----YHVLVLSWRSGSSSFV 57
 DB 21 LARGMRLPFRSSTVMSLVMGTGLVFLVSNQVSSPAGLGERVHVLVLSWRSGSSSFV 80
 DB 58 GOLFQHPDYFLMEPRAMHVMFTFTSTAMKILMAVRDLRSVFLCDMSVFDAYMNPGR 117
 DB 81 GOLFQHPDYFLMEPRAMHVMFTFTSTAMKILMAVRDLRSVFLCDMSVFDAYL-PMRR 139
 QY 118 KOSLFEWESRALCSAPVCDPEFAHEISPKHCKILCGQOPDMVEKACRSHGFVYLKE 177
 DB 140 NISLFEWAVSRALCSPPVCEAFARNTSSEVCKPLCATRFGLAQECSSYSHVYLKE 199
 QY 178 VRFSLQALYPLTDPSTLHVHVLVRDPAVRSRREHTTIELMVDSHVILGOHETIKE 237
 DB 200 VRFSLQALYPLTDPSTLHVHVLVRDPAVRSRREHTTIELMVDSHVILGOHETIKE 258
 QY 238 EDQPYAMKTIKSOVDIYKAIQTLPEALQORLYFLRVEDLVRAPLAQTRLYKVGDF 297
 DB 259 ADPLRVNVEYCSHRVIAEALHMKRPPFLQDRYRLVRYEDLARDLTVIRELYAFTGIGL 318
 QY 298 LPHLOTWYNYVRGKGMG--OHAFTNARNALNVSQAMRWSLPIYEKVSQLODACEAMD 355
 DB 319 TPQLDTWHTNTHGSGPCARRAEFKTTSKDALVSQAMRHLLPFAKIRVQELCGALQL 378
 QY 356 LGYLQVRSQOEGNSLDL 375
 DB 379 LGYRSVHSELRODLSLDL 398

RESULT 8
 AAY72638
 AAY72638 standard; Protein; 395 AA.

AAV72638;

02-MAY-2001 (first entry)

Mouse glycosyl sulfotransferase-4 (GST-4).

Mouse; glycosyl sulfotransferase-4; GST-4; immunosuppressive;
 selectin binding inhibitor; gene therapy; inflammation;
 systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
 polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
 glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;
 Hashimoto's disease; Grave's disease; hypoparathyroidism; anemia;
 demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
 myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
 asthma; hypersensitivity; rheumatic fever; tissue rejection;
 chromosome 8E1.

Mus musculus.

MO200106015-A1.

25-JAN-2001.

19-JUL-2000; 2000MO-US19741.

20-JUL-1999; 99US-0144694.

PR 13-JUL-2000; 2000US-0593828.

XX (REGC) UNIV CALIFORNIA.

XX Rosen SD, Lee JK, Hemmerich S;

XX WPI: 2001-138471/14.

DR N-PSDB; AAD02696.

PT New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
 XX diagnostic and therapeutic agent screening applications -

PS Claim 3; Fig 2; 128pp; English.

CC The present sequence is mouse glycosyl sulfotransferase-4 (GST-4). GST-4
 CC gene is found on chromosome 8E1.
 CC GST is a type 2 membrane protein useful for inhibiting a binding event
 CC between a selectin and a selectin ligand, which comprises contacting the
 CC selectin with a non-sulphated selectin ligand, GST and a small molecular
 CC agent that inhibits the sulphation activity of GST. GST is also useful
 CC in inhibiting a selectin mediated binding event. GST is useful in gene
 CC therapy to treat disorders such as acute or chronic inflammation,
 CC systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
 CC nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
 CC glomerulonephritis, myasthenia gravis, Sjogren's syndrome, pernicious
 CC anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
 CC dermatitis, myocarditis, regional enteritis, adult respiratory distress
 CC syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
 CC bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
 CC during transplantation.

SQ Sequence 395 AA;

Query Match 48.2%; Score 984; DB 22; Length 395;
 Best Local Similarity 54.9%; Pred. No. 7.3e-96;
 Matches 207; Conservative 44; Mismatches 114; Indels 12; Gaps 6;

QY 8 RLMLFLGSOVYVALFIHMSVHRHLSORESRP-----YHVLVLSWRSGSSSFVGLF 61
 DB 2 RLPRF-SSTVMSLVMGTGLVFLVSNQVSSPAGLGERVHVLVLSWRSGSSSFVGLF 60
 QY 62 GHPDYFLMEPRAMHVMFTFTSTAMKILMAVRDLRSVFLCDMSVFDAYMNPGRKSS 121
 DB 61 GHPDYFLMEPRAMHVMFTFTSTAMKILMAVRDLRSVFLCDMSVFDAYL-PMRRNISD 119
 QY 122 LEWESRALCSAPVCDPEFAHEISPKHCKILCGQOPDMVEKACRSHGFVYLKEVFL 181
 DB 120 LEWESRALCSAPVCDPEFAHEISPKHCKILCGQOPDMVEKACRSHGFVYLKEVFL 179
 QY 182 SLQALYPLTDPSTLHVHVLVRDPAVRSRREHTTIELMVDSHVILGOHETIKEEDP 241
 DB 180 NLQVLYPLTDPSTLHVHVLVRDPAVRSRREHTTIELMVDSHVILGOHETIKEEDP 238
 QY 242 YRAMKTIKSOVDIYKAIQTLPEALQORLYFLRVEDLVRAPLAQTRLYKVGDFLPH 300
 DB 239 LRVNVEYCSHRVIAEALHMKRPPFLQDRYRLVRYEDLARDLTVIRELYAFTGIGL 298
 QY 301 LOTWYNYVRGKGMG--OHAFTNARNALNVSQAMRWSLPIYEKVSQLODACEAMDLYG 358
 DB 299 LOTWYNYVRGKGMG--OHAFTNARNALNVSQAMRWSLPIYEKVSQLODACEAMDLYG 358
 QY 359 LGYRSVHSELRODLSLDL 375
 DB 359 RSVHSELRODLSLDL 375

RESULT 9

AAU11275

AAU11275 standard; Protein; 395 AA.

AAU11275;

[illegible]

Dd		120	LFGWNASRALCSPPVEAARNGNISSEBCKRLCATRPFGLAQEACSSYSHVVLKEVRFF	179
Oy		182	SLQALYPLLTDPSPSLNLHVHVLVDPRPAVRSREHTTEILMWDHSIVLGOHLETKREDOP	2411
Dd		180	NQLVPLELSDPALNRIYHVLVDPRPAVRSREQAKALARDNGIVLGCTN-GTWVEADPR	238
Oy		242	YYAMKTIICKSQNDVIYK-AIQOTLPBALQQRYLFLEFREDLVRRPPLAQOTRLRYKFVGDLPH	3000
Dd		239	LRVVNEGCRSHVAIAEAALHKPPPELDQDXYRLVREDLARPOLYVIRELVAFTGLGLTPQ	298
Oy		301	LQTWVYNVTGSKMG-MG-OHAFTTNARNALNWSOAWRMSLPYEKYSQLODACGEAMDLLGY	358
Dd		299	LQWMINITHIGSGPGARRAEAFRTTSDALSVSQAARHRTLPRAKIRRODELGGMALDGLGY	358
Oy		359	LQVRSQEQEGCNLSLDLL	375
Dd		359	RSVHSELQGRDLSLDLL	375
 RESULT 10 ABB81555 ID ABB81555 standard; Protein; 395 AA. XX ABB81555; AC XX DT 05-SEP-2002 (first entry) Xx Consensus N-acetylglucosamine-6-sulfotransferase SEQ ID NO:3. De De Human: N-acetylglucosamine-6-sulfotransferase; enzyme: G1CNA6ST; Kw corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD; Kk ophthalmological. Xx Homo sapiens. Os Mus musculus. Os Synthetic.				
Xx	Key	Location/Qualifiers		
Ft	Misc-difference	10 /label= Ala, Thr, Val		
Ft	Misc-difference	13 /label= Ala, Val, Ser		
Ft	Misc-difference	20 /label= Phe, Cys, Gly		
Ft	Misc-difference	39 /label= Ala, Asp, Glu		
Ft	Misc-difference	96 /label= Val, Met, Ile		
Ft	Misc-difference	142 /label= Ala, Thr, Asn		
Ft	Misc-difference	147 /label= Ala, Asp, Glu		
Ft	Misc-difference	159 /label= Thr, Ser, Gly		
Ft	Misc-difference	238 /label= Gly, His, Arg		
Ft	Misc-difference	294 /label= Ser, Thr, Gly		
Ft	Misc-difference	371 /label= Ala, Thr, Ser		
Ft	Misc-difference	380 /label= Leu, Pro, Met		
Ft	Misc-difference	382 /label= Gly, His, Ser		
Ft	Misc-difference	384 /label= Thr, Ser, Lys		
Ft	Misc-difference	390 /label= Ala, Glu		
Ft	Misc-difference	391 /label= Ser, Lys		
Ft	Misc-difference	392 /label= His, Glu		
Ft	Misc-difference	394		

QY	111	YMNQPRROSSLTFQWQEQSRALCSAPVCDPEFPAHEISSPKCHKCLGQOPEFDMWEKCRSH	170
Db	110	YL-FMRRLMSDLDFQWNAVRALCSPPACGASFPFGALISSEAVCKPCLCARQSFLLARECRST	168
QY	171	GFVVLKEVREYSLQALVPLLTDPSLNLAVVHLVRDRAVERSRSEHTTITELMVDSHVILGO	230
Db	169	SHVVLKEVREYFNLQVLYPLDPSLPALNLRIVHLVRDPRVALRSREQTAKALARNGVILGT	228
QY	231	HLFEIKEDOPFYAKKTIKCSQOVDLYVAKIQT-L-PEALQORYFLFLEEDLVRAPIAQTTR	288
Db	229	N-GTWVEADPGELRAVREYCRSHVRIAE-A-TLKPPFLGRLKRLVYFEDLAREPPLAEIRA	286
QY	289	LYKEVGLDLEPLHLQTVVYNVTRGKMG--QHAFTHTARNALNVSQAMRSLPYEKYSOLO	346
Db	287	LYAFGLSLTDLQLEAMINHITHGSGPGARERAEFKTSSRALNVSQAMRHRLPFAKTRVQ	346
QY	347	DACGEAMDLGLGYVRSOQEOGNLSIDL	375
Db	347	ELCAGALDLDLGRFYISEDEQRNLALDLY	375

XX Sequence 390 AA; 46.9%; Score 958.5; DB 22; Length 390;
 SQ Query Match Best Local Similarity 55.6%; Pred. No. 3.8e-93;
 Matches 193; Conservative 42; Mismatches 87; Indels 25; Gaps 6;

QY 42 VHVIVLSMRSGSFFVGLFQGHDPVFLMEPAHVMWTFSTAMKLMHVRDLRSVF 101
 DB 42 VHVIVLSMRSGSFFVGLFQGHDPVFLMEPAHVMWTFSTAMKLMHVRDLRSVF 101
 QY 102 LCDMSVFDAYMNPGRKQSLFQWESRALCSAPYCDFFPAHEISSPKHCKLGGQOPFD 161
 DB 102 LCDMDVFDAYM-POSRLMSAFENWATSRALCSAPYCDFFPAHEISSPKHCKLGGQOPFD 160
 QY 162 MVEKACRSHGFVVLKEVRFSLQALYPLTDPSLNLHVHVLVDRPRAVRSREHTTIELM 221
 DB 161 LAREACRSYSHVVLKEVRFSLQALYPLTDPSLNLHVHVLVDRPRAVRSREHTTIELM 220
 QY 222 VDSHIVLG-----QHLETIKEEDPYAMKICKSOVDIYKAIOTL--PEALQORY 270
 DB 221 RDNGIVLTGNCKWEADPHLRLIRE-----VCRSHVRIAEA-ATLKPPLPLRGRY 269
 QY 271 LFLRYEDLVRAPIAQTTRLYKFEVGLDFLPHLQWVYVNTYRGKMGQ--HAFTNARALN 328
 DB 270 RLVRFEEDLAREPLAEIRALYFTGLTLPLQLEAMIHNTHTSGIGKPIEAFHTSSRNARN 329
 QY 329 VSQAWRMSLPYEKYSQDADGCEAMDILGYIQVNSQOQNLSDLL 375
 DB 330 VSQAWRHALLPFTKILRVQEVCGALQLLGYRPPVYSAQQORDLTLDLV 376

RESULT 15
 ABB81556
 ID ABB81556 standard; Protein: 390 AA.
 AC ABB81556;
 XX 05-SEP-2002 (first entry)
 DT 05-SEP-2002 (first entry)
 XX Human intestinal N-acetylglucosamine-6-sulfotransferase SEQ ID NO:4.
 DE Human intestinal N-acetylglucosamine-6-sulfotransferase; enzyme: GLCNAC6ST;
 KW corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
 KW ophthalmological.
 XX Homo sapiens.
 OS Homo sapiens.
 PN US2002061562-A1.
 XX 23-MAY-2002.
 XX 09-AUG-2001; 2001US-0927602.
 XX 11-AUG-2000; 2000US-325773P.
 XX (FUKU/) FUKUDA M N.
 XX (AKAM/) AKAMA T O.
 XX Fukuuda MN, Akama TO;
 XX WPI: 2002-507643/54.
 DR New nucleic acid encoding corneal
 PT N-acetylglucosamine-6-sulfotransferase, useful for treatment,
 PT monitoring and diagnosis of macular corneal dystrophy
 XX Example 5; Fig 2A-B; 69pp; English.
 CC The present invention describes human corneal
 CC N-acetylglucosamine-6-sulfotransferase (I), which is able to catalyze
 CC sulfation of keratan sulfate (KS). Also described is a method for
 CC monitoring the effect of treatments for macular corneal dystrophy (MCD),

CC and detecting susceptibility to MCD. (I) is located to chromosome 16q22,
 CC and has ophthalmological activity. (I) can be used to treat or prevent
 CC macular corneal dystrophy types I or II. (I) makes possible treatment
 CC of MCD without requiring keratoplasty or keratectomy. The present
 CC sequence represents human intestinal N-acetylglucosamine-6-
 CC sulfotransferase, which is given in comparison with (I) in the
 CC exemplification of the present invention.

XX Sequence 390 AA; 46.9%; Score 958.5; DB 23; Length 390;
 SQ Query Match Best Local Similarity 55.6%; Pred. No. 3.8e-93;
 Matches 193; Conservative 42; Mismatches 87; Indels 25; Gaps 6;

QY 42 VHVIVLSMRSGSFFVGLFQGHDPVFLMEPAHVMWTFSTAMKLMHVRDLRSVF 101
 DB 42 VHVIVLSMRSGSFFVGLFQGHDPVFLMEPAHVMWTFSTAMKLMHVRDLRSVF 101
 QY 102 LCDMSVFDAYMNPGRKQSLFQWESRALCSAPYCDFFPAHEISSPKHCKLGGQOPFD 161
 DB 102 LCDMDVFDAYM-POSRLMSAFENWATSRALCSAPYCDFFPAHEISSPKHCKLGGQOPFD 160
 QY 162 MVEKACRSHGFVVLKEVRFSLQALYPLTDPSLNLHVHVLVDRPRAVRSREHTTIELM 221
 DB 161 LAREACRSYSHVVLKEVRFSLQALYPLTDPSLNLHVHVLVDRPRAVRSREHTTIELM 220
 QY 222 VDSHIVLG-----QHLETIKEEDPYAMKICKSOVDIYKAIOTL--PEALQORY 270
 DB 221 RDNGIVLTGNCKWEADPHLRLIRE-----VCRSHVRIAEA-ATLKPPLPLRGRY 269
 QY 271 LFLRYEDLVRAPIAQTTRLYKFEVGLDFLPHLQWVYVNTYRGKMGQ--HAFTNARALN 328
 DB 270 RLVRFEEDLAREPLAEIRALYFTGLTLPLQLEAMIHNTHTSGIGKPIEAFHTSSRNARN 329
 QY 329 VSQAWRMSLPYEKYSQDADGCEAMDILGYIQVNSQOQNLSDLL 375
 DB 330 VSQAWRHALLPFTKILRVQEVCGALQLLGYRPPVYSAQQORDLTLDLV 376

Search completed: August 9, 2003, 16:42:40
 Job time : 40.6021 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: August 9, 2003, 16:40:34 ; Search time 15.0388 Seconds
(without alignments)
2481.152 Million cell updates/sec

Title: US-09-645-078-4

Perfect score: 2042

Sequence: 1 MMLKKRLLMFLGSGVIV.....NLSDLLSSHLIGQVREG 388

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 9616682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: p1r1:*
2: p1r2:*
3: p1r3:*
4: p1r4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	593	29.0	486	2 JC7351	N-acetylglucosamin
2	591.5	29.0	484	2 JC7350	N-acetylglucosamin
3	574.5	28.1	484	2 JE0261	N-acetylglucosamin
4	496.5	24.3	458	2 A57397	chondroitin 6-sulf
5	134.5	6.6	388	2 G70729	hypothetical prote
6	124	6.1	307	2 E95934	probable enzyme, C
7	101.5	5.0	4273	2 C69679	polyketide synthas
8	98.5	4.8	3187	2 JC5837	364k Golgi complex
9	97.5	4.8	291	2 F38626	Me240 homolog F10-
10	94.5	4.6	513	2 F90866	transcription regu
11	93	4.6	346	2 E86319	probable cyclase -
12	93	4.6	726	2 T35865	transcription regu
13	92.5	4.5	292	2 AD0362	probable lysr-fami
14	92.5	4.5	513	1 RGE043	transcription regu
15	92.5	4.5	1495	2 S61023	hypothetical prote
16	92	4.5	812	2 D84964	transcription-repa
17	91.5	4.5	338	2 A96769	protein flavonol s
18	91.5	4.5	513	2 C85752	hypothetical prote
19	91.5	4.5	1169	2 F97872	transcription-repa
20	91	4.5	2354	2 T13388	mei-41 protein - I
21	90.5	4.4	1169	2 F95000	transcription-repa
22	89	4.4	361	2 T41176	hypothetical prote
23	89	4.4	910	2 E89918	2-oxoglutarate deh
24	88	4.3	469	2 D84857	hypothetical prote
25	88	4.3	905	2 B71562	probable salt-indu
26	87	4.3	822	2 T01622	probable salt-indu
27	86.5	4.2	467	2 F83651	hypothetical prote
28	86.5	4.2	664	2 T48258	kinesin-like prote
29	86.5	4.2	892	2 T27005	hypothetical prote

30	86.5	4.2	1116	2 H97970	type I site-specif
31	85.5	4.2	974	2 A44484	transcription regu
32	85.5	4.2	1091	2 B95103	hypothetical prote
33	85	4.2	344	2 G96599	protein F14J16.8
34	85	4.2	844	2 S37794	aminopeptidase Y9c
35	84	4.1	995	2 A84014	SNF2 helicase BH29
36	84	4.1	1293	2 T27886	hypothetical prote
37	84	4.1	1813	2 T19295	hypothetical prote
38	83.5	4.1	424	2 A71476	probable protopor
39	83.5	4.1	623	2 G81420	hypothetical prote
40	83.5	4.1	960	1 OX12	phosphoenolpyruvat
41	83	4.1	398	2 C86749	hypothetical prote
42	83	4.1	513	2 AE0659	transcription regu
43	83	4.1	655	2 G87306	threonyl-tRNA synt
44	83	4.1	776	2 T20896	hypothetical prote
45	83	4.1	1568	2 T41013	hypothetical prote

ALIGNMENTS

RESULT 1

JC7351
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - human
C:Species: Homo sapiens (man)
C:Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
C:Accession: JC7351
R:Uchimura, K.; Fasakhany, F.; Kadomatsu, K.; Matsukawa, T.; Yamakawa, T.; Kurosawa, Blochem. Biophys. Res. Commun. 274, 291-296, 2000
A:Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of A:Reference number: JC7350; MUID:20374462; PMID:10913333
A:Accession: JC7351
A:Molecule type: mRNA
A:Residues: 1-486 <UCH>
A:Cross-references: DDBJ:AB040711
C:Comment: This enzyme, having a type II transmembrane topology and N-linked glycosyl sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand C:Keywords: glycolysis; sulfate transport; sulfotransferase; transmembrane protein

Query Match 29.08; Score 593; DB 2; Length 486;
Best Local Similarity 37.78; Pred. No. 4.6e-44;
Matches 146; Conservative 51; Mismatches 136; Indels 54; Gaps 10;

OY	38	SRPRVHYLVSSWSSGSPVGLFGOHDPVYLMPEAMHWMTTSSPAKMLHVAVDL	97
DB	98	SREKQHTYHATVMTGSSFLGELNQHDPVYLTPEMHMLQALYPGDAESLQALDML	157
OY	98	RSVFCDMSYFDAYMNGP-PRKQ-----SSLFQWEDSRALCSAPVCDFFPAHE---	144
DB	158	RSLEKDFSVLRLYAPGDPAPARAPDTANLTALFMRRTNKVICSPLCGAPARAAY	217
OY	145	-ISSPKCKLLCGQGPDPDMEKACRSHGCVYLVKAVRSLQALYPLTDSLNHVVHLV	203
DB	218	GLVEDTACRSCSPVAIRALEAECKKYPVVVIXKIDVRLDGLVPLLRDPLNKKVOLF	277
OY	204	RDPRAVFRSREHTTEILMDS-----HYLVGOH-----LETIKEE	238
DB	278	RDPRAVNSRKSQGLRESIOVLRTRQGRDHRHVLHAGVGARGGSRALPAAPRA	337
OY	239	DOPYR---AKIICKSOVDIVKAIOTLPEALQRYLFLRYEDLVRAPLAOTTRLYKFGVL	295
DB	338	D--FELGALVEICEALRDLFLFARGAPAMLRRTYLRLEDLVKRGARQLRLRRSGL	395
OY	296	DFLPHLOTWYVNTYTRGKMG-QIAEFHNARNALNVSAWKMSLPEYVSQADAGGAMD	354
DB	396	RALNALDPAFLNMTRGAAVYAGADRPFLHSARDAEVAHMERLSRQVROVEACAPAMR	455
OY	355	ILGYLGVRSQOE-----OGNLSLDL	374
DB	456	ILAY--PRSGEBGDAEQPREGTEPLEM	480

RESULT 2

JC7350
N-acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) [validated] - mouse

C.Species: Mus musculus (house mouse)
C.Date: 08-Sep-2000 #sequence_revision 08-Sep-2000 #text_change 31-Dec-2000
A.Accession: JC7350
R.Uchiyama, K.; Fasaki, H., T.; Matsukawa, Y.; Yamakawa, T.; Kurosawa, N.
Biochem. Biophys. Res. Commun. 274, 291-296, 2000
A.Title: Diversity of N-acetylglucosamine-6-O-sulfotransferases: Molecular cloning of a
A.Reference number: JC7350; MUID:20374462; PMID:10913333
A.Accession: JC7350
A.Molecule type: mRNA
A.Residues: 1-484 <UCH>
A.Cross-references: DDBJ:ABO40710
C.Comment: This enzyme, having a type II transmembrane topology and N-linked glycosylation sulfate linkage formation. This enzyme is involved in synthesis of L-selectin ligand in C-keywords: glycopolis; sulfate transport; sulfotransferase; transmembrane protein.

Query Match 28.1%; Score 574.5; DB 2; Length 484;
Best Local Similarity 35.0%, Pred. No. 6.2e+44;
Matches 136; Conservative 49; Mismatches 111; Indels 41; Gaps 6;

OY 43 HTVLSSRRGSGSVGGGLFGOHDPVFLYEPMHVMWTFSTAWKLMHAYRLRSFVL 102
Dd 101 HIYYHAATWRGTGSSEIGELFNQHPDFYLPEMHLMOALYPGDASESLGALLRDMLRSIFL 160
OY 103 CDMSVFPAVNMPG-PFRQ-----SLSFMQGSRALCAPVCDEFPPANE----ISSP 148
Db 161 CDSEVLRILYAPGPGEAPDSANLTTPAMFERMTNTNYCSPLCPAAPRARADVLED 220
OY 149 KHCKILCGQPFPDMVEAKRGSHGFNVLEKEAFYSIALYLPDLPSINTLVHYHLVRPRA 208
Db 221 KACSTCPSPVALRALAEBCRYPYVVYIKDVALLDGVLVPILLRPGLNLKVQLFRPRA 280
OY 209 VFSREERTTIELAMDShIVL-----GHLEFT-----KEEDOPY 242
Db 281 VNHSRLKSROGLRESIOVLTQRGDHFHVLLAHGVADARPQAARLBPAPADEFLT 340
OY 243 YAMIICKSOVDIYKALOTLPDALOORYLFERYEDELVARPIAQTRTLKYFGLDFPHIQ 302
Db 341 SALEVICETAMLRIIDLPFRGARPARLTRIRLYRIDVLDWMOPAOURLRLRESGLFTLAJD 400
OY 303 TWYNVTNRKGKMG-QHAFFHTARNALNAVSAQRSLSLEYEKVSQOLDAAGEMDLIIGY 358
Db 401 AFAENMTGRGAAGYADDREFHLASARDAREAVHWHRRLSQEQRYQTETCAPRMRLAY 457

RESULT 3
JE0261
N-Acetylglucosamine-6-O-sulfotransferase (EC 2.8.2.-) - human
C.Species: Homo sapiens (man)
C.Date: 05-Feb-1999 #sequence_revision 05-Feb-1999 #text_change 01-Mar-2002
A.Accession: JE0261
R.Uchiyama, K.; Muramatsu, H.; Kaname, T.; Ogawa, H.; Yamakawa, T.; Fan, Q.W.; Mitsuoaka,
J. Biochem. 124, 670-678, 1998
A>Title: Human N-Acetylglucosamine-6-O-sulfotransferase involved in the biosynthesis of
A.Reference number: JE0261; MUID:98391845; PMID:9722682
A.Accession: JE0261
A.Molecule type: mRNA
A.Residues: 1-484 <UCH>
A.Cross-references: DDBJ:AB014679
C.Comment: This protein catalyzes the transfer of sulfur of sulfate from 3'-phosphoadenine
C.Superfamily: chondroitin 6-sulfotransferase
C.Keywords: sulfoltransferase

Query Match 28.1%; Score 574.5; DB 2; Length 484;
Best Local Similarity 35.0%, Pred. No. 1.9e+42;
Matches 130; Conservative 65; Mismatches 121; Indels 55; Gaps 9;

OY 46 VLSSMRGSSSFVGGLFGCHPDVFLYEPMHVMWTPTSTAWKLHMAYRDLIRSFIICDM 105
Dd 123 VTITWRGSSPFEGGLFNONEEVFLYPEVWHWKLYPGDAVSLQGAARDMLSLTYRCDL 182

QY	SVFPAVYNNPGRKRS	----	SLFQWQSQSLASAPCDFFPHE	-----	ISSPKHCKLQCG	156
QY	106	SVFPAVYNNPGRKRS	----	SLFQWQSQSLASAPCDFFPHE	-----	ISSPKHCKLQCG 156
Db	183	SVFQSLYSAGSGGRNLTLLGIFGAATKNVYVSSPFC	----	PAYRKEVVALVDYDRVK	-KCP	238
QY	157	QQPFDMWEKAKRSHGFEVVLKEVEVRLSLQALVPLLTDPSSLNTHVHLVBDPRAVRSREHT	216			
Db	239	QQLARFECECRKRYTLLKGVAVBDVAVLPLMLDPALDLKVLHVLBDPRAVASSR	---	295		
QY	217	TIELMVDSHIYLGQHLERIKED	-----	QPYV	-----	AKI 247
Db	296	-----	IRSHGRLRSLOQVRSRDRDPRAHMRPLLEAGHKLGAKKEGVSGRADYHALGAMEV	351		
QY	248	ICKSQVDIVAKIQTFL	-----	PEALQORRYLELRYEDLVRAPLAOTTRLYKKGVGIDFLPHLOTW	304	
Db	352	ICNS	----	MAKTLQALQAPRPMQLQGHLYLVRRREDLVGEVKTLLRRYYDVGVLVSPENQF	408	
QY	305	VYNTTRGKMGQHAFTNARNAALNVSQAMRNSLPREKYSQLODACGEMDLGTYQVRSQ	364			
Db	409	ALNMTSGSGSSSKPFVVSARNATQAMAMRTALTFOQIKQVEEFCYQPMALVGERVNSP	468			
QY	365	QEGNLTSLDL	375			
Db	469	EEVKDLSTKL	479			
RESULT 4						
A57397						
chondrotilin 6-sulfoltransferase (EC 2.8.2.17) precursor - chicken						
C:Species: Gallus gallus (chicken)						
C:Date: 23-Feb-1996 #sequence_revision 23-Feb-1996 #text_change 20-Jun-2000						
C:Accession: A57397						
R:Fukuta, M.; Uchamura, K.; Nakashima, K.; Kato, M.; Kimata, K.; Shinomura, T.;						
J. Biol. Chem. 270, 18575-18580, 1995						
A:Title: Molecular cloning and expression of chick chondrocyte chondrotilin 6-sul-						
A:Reference number: A57397; MUID:95355490; PMID:7629189						
A:Accession: A57397						
A:Status: preliminary						
A:Molecule type: mRNA						
A:Residues: 1-458 <FUK>						
A:Cross-references: GB:D9915; NID:9971262; PIDN:BA08655.1; PID:9971263						
C:Superfamily: chondrotilin 6-sulfoltransferase						
C:Keywords: sulfotransferase						
Query Match						
Best Local Similarity 32.7%; Score 496.5; DB 2; Length 458;						
Matches 118; Conservative 67; Mismatches 133; Indels 43; Gaps 10;						
QY	37	ESRRPVNHYLVSSWRSGSPYQQLFGQHPDYFLMEPRAHYMMTFT	-----	SSTANKLMA	92	
Db	110	EPRR--HYLLMATYRTGSSFGGEFFNOGINIFYLEPLMHIERTVTEFPBGANAVASALY	167			
QY	93	VRDLRSVFLCDMSVFDAVYNNPGRKQ--SLFQWQSQSLASAPVCDFFPA-HEISSPK	149			
Db	168	YRDVLQQLLDLDLILEFISPAPEEHITALLFRGSSHSICEEVCY--PSLKVFEEKY	225			
QY	150	HCK-LLCQGFDMWEKAKRSHGFEVVLKEVEVRLSLQALVPLLTDPSSLNTHVHLVBDPRA	208			
Db	226	HCKNRRCGRPLNITLAAEKCRKQNHAKLTAVRIQGEFLQPLAEDPRDLRIIQLVBDRA	285			
QY	209	VFRSREHTTIELMVDSHIYLGQHLERIKEDQPYTAMTIICKSQVDIYKAIQ	-----	260		
Db	286	VLVSR-----	WAFSGKYESMK	-----	WAGEAPLQDEDEQVRKNCESIRL	329
QY	261	-----	TLPEALQORYLELRYEDLVRAPLAOTTRLYKKGVGIDFLPHLOTWYVNTYRGKM	314		
Db	330	SAELGLRQPRMLRGKMYLVRYEDVAPARLRALETRYRAGIHPPIQVEWETIRANTQAP-Q	388			
QY	315	GQHAFTNARNAALNVSQAMRNSLPREKYSQLODACGEMDLGTYQVRSQ	364			
Db	389	DSNGLIYSQKNSSQPFEMKRSIFPKLQVYQDACEPRAMRLFGYKLASSADELTNRSL	448			
QY	375	L	375			

Db 449 L 449

RESULT 5

G70729 hypothetical protein RV2267c - Mycobacterium tuberculosis (strain H37Rv)

C:Species: Mycobacterium tuberculosis

C:Date: 17-Jul-1998 #sequence_revision 17-Jul-1998 #text_change 22-Oct-1999

C:Accession: G70729

R:Colet, S.T.; Brosch, R.; Parkhill, J.; Garnier, T.; Churcher, C.; Harris, D.; Gordon, S.

R:Connot, R.; Davies, R.; Devlin, K.; Feltham, T.; Gentles, S.; Hamlin, N.; Holtroyd, S.

R:Rajendram, M.A.; Rogers, J.; Rutter, S.; Seeger, K.; Skelton, S.; Squares, S.

R:Nature 393, 537-544, 1998

A:Authors: Squares, R.; Sulston, J.E.; Taylor, K.; Whitehead, S.; Barrell, B.G.

A:Title: Deciphering the biology of Mycobacterium tuberculosis from the complete genome

A:Reference number: A70500; MUID:96295987; PMID:9634230

A:Accession: G70729

A:Status: preliminary; nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-388 <COL>

A:Cross-references: GB:277163; GB:AL123456; NID:93261610; PIDN:CAB00968.1; PID:e255072;

A:Experimental source: strain H37Rv

C:Genetics:

A:Gene: RV2267c

Query Match 6.6%; Score 134.5; DB 2; Length 388;

Best Local Similarity 21.8%; Pred. No. 0.00059;

Matches 77; Conservative 59; Mismatches 116; Indels 101; Gaps 20;

QY 44 VLVLSWRSGSSFFVGLF---GQH-----PDVFYLMEPAMHV-WMTFTSSTWAKL 89

Db 83 IFYIGHWRTGTTLLHLLVVDHRTGPTGCECLAPHHFLTE--WPAHYEELVSK----- 136

QY 90 HMAVRDLRSVFLCDMSVFDAVYMPGPKQSLFQWESRALCSPVCDFFPAHEISSPK 149

Db 137 HRADNN-----DLSLH-----HPQDEFVW-CMOGLPSPLYLTAEPNR---PPQ 177

QY 150 HCKLLCGQPFDPWEKACRSHGFVLEKVEFSLQALY-----PLTDPSLNHV--- 199

Db 178 Y-----EYKLDQVAPRE--LEIKWRTLEFRVQVYFRKRYVILKNPHSRIVKL 229

QY 200 -----VHLRDPRAVFRSREHTTIELVDSHIVLGQHLTIKEEDQPYAMKIKS 251

Db 230 EVFQAKFIHVRDPPVYVPS---TILHKAFLYRIGLQPTFDGDK-----VYST 279

QY 252 QVNDVKAIGTLPEALQ-QRYLFLEYEDLVAPLAQTRIKYFGVLD---FLPHLQTVWY 306

Db 280 YVDLYRRLDREGRLEVDPTRFYELEEDLIGDPGQLRXYOHLGIDGFECLPRLRQY-- 337

QY 307 NVTRGKMGQHA-FHTNARNALNVSAQMRMSLPYEKVSQLODAGCEAMDILGY 358

Db 338 -----LADHADYKINS-----YGLIVEQRAIVDEHGEIIDRIGY 372

RESULT 6

E95934 probable enzyme, C-terminal domain similar to sulfotransferase protein [imported] - Sinc

C:Species: Shorhizobium meliloti

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: E95934

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernat

Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1.683-kb pSymb megaplasmid from the N2-fixing endo

A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: E95934

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-307 <KOR>

A:Cross-references: GB:AL591985; PIDN:CAC49141.1; PID:915140626; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSymb

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler,

Pela, D.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.

L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelau

hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh,

A:Title: The composite genome of the legume symbiont Shorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: Smb21237

A:Genome: plasmid

Query Match 6.1%; Score 124; DB 2; Length 307;

Best Local Similarity 21.3%; Pred. No. 0.0036;

Matches 79; Conservative 54; Mismatches 120; Indels 118; Gaps 19;

QY 40 RPVHVLVSSW-RSGSSFFVGLFQGHDPVF---YLMEPAMHV-----MTF 81

Db 4 QPVIAYIAGYGRSGSTLIDIALGQHAHVAGCETTSITRHVMHNEACAGNAIRDCSF 63

QY 82 TSSP--AMK-----LHMAVRDLRSVFLCDMSVFDAVYMPGPKQ 119

Db 64 WSSVREWSGQDPGLMEYCALQOKREGLSMTR-LTSGH-----GLGKQ 108

QY 120 SLPQWESRALCSPVCDFFPAHEISSPKHCKLLCGQPFDPWEKACRSHGFVLEKVR 179

Db 109 FSLYTLHTRKLFSAHQSS-----GRQ---VYDSSKLPG-----R 141

QY 180 FLSLQALYPLTDSNLHVVHLYRDPRAVFRSREHTTIELVDSHIVLGQHLT---IK 236

Db 142 AMAVAOI-----GQIDRVYTLVRDGVGAVMS-----LLKGERPAKSGQ 182

QY 237 EEDQPYAMKRIK-SQVDYKAIOTLPAL-QOYLFLEYEDLVAPLAQTRIKYFGV 294

Db 183 KEIKPKSVFRALMNSVNL--AVEYLSRKLGSEKVMKVRKEDASDPVAMQIGTFLE 240

QY 295 LDPLPHLQTVWYNTVRGKMGQ-----HAETNARNALNVSAQMRMSLPYEKVSQLODA 348

Db 241 LD-LSQVGTSLDN--GEAMGPHQVAGNRLRNANASIALNDETFWTRMPAQOVSPORL 296

QY 349 CGEAMDILGYL 359

Db 297 GGMMLRRYGYL 307

RESULT 7

C69679 polyketide synthase pksM - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 03-Nov-2000

C:Accession: C69679

R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Ber

C.; Bron, S.; Brouillet, S.; Brusch, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.;

A.; Ehrlich, S.D.; Emerson, P.T.; Entian, K.D.; Erttington, J.; Fabel, C.; Ferrari,

Nature 390, 249-256, 1997

A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gal

lechi, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holsappel, S.; Hosono, S.; Hullo, G

koetter, P.; Konigstein, G.; Krogh, S.; Kumano, M.; Kurita, K.; Lapidus, A.; Lardino

A:Authors: Lander, U.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Mau

y, M.; Ogawa, K.; Ogatawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portete

Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadale, Y.; Sato, T.; Scanl

A:Authors: Schleich, S.; Schroeter, R.; Scoffone, F.; Sekiguchi, J.; Sekowska, A.; Se

akeuchi, M.; Tamakoshi, A.; Tanaka, T.; Terpilera, P.; Tognoni, A.; Tosato, V.; Uchida

T.; Winters, P.; Wipat, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida

A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, H.; Danchin, A.

A:Title: The complete genome sequence of the Gram-positive bacterium Bacillus subtilis

A:Reference number: A93580; MUID:96044033; PMID:9384377

A:Accession: C69679

A:Status: preliminary; nucleic acid sequence not shown; translation not shown.

A:Molecule type: DNA

A:Residues: 1-4273 <RUN>

A:Cross-references: GB:299113; GB:AL009126; NID:92634090; PIDN:CAB13603.1; PID:926341

A:Experimental source: strain 168

C:Genetics:

A:Gene: pksM
 C:Superfamily: Bacillus subtilis polyketide synthase pksM; 3-oxoacyl-[acyl-carrier-prote
 C:Keywords: carrier protein
 F:293-363/Domain: acyl carrier protein homology <ACPI>
 F:144-829/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OASI>
 F:2340-2732/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS>
 F:3145-3320/Domain: short-chain alcohol dehydrogenase homology <SADI>
 F:3350-3842/Domain: 3-oxoacyl-[acyl-carrier-protein] synthase I homology <OAS3>
 F:4138-4208/Domain: acyl carrier protein homology <ACP>

Query Match 5.0%; Score 101.5; DB 2; Length 4273;
 Best Local Similarity 22.3%; Pred. No. 11;
 Matches 59; Conservative 44; Mismatches 99; Indels 63; Gaps 14;

153 LILGQGFDPVWEKACRSHGFVLEKREPLSLQALYPLTTPS-----LNLHVHLYVLRPRA 208
 DB LILGFLAYIDLYQVQVEHGY-AYQELKLNLTIFVPLADESYDIALFTIH----- 74
 209 VFRSHHTIELMVSHVILGHTETKE-----EQPYAMKIIICKSQVDYK 257
 DB --SEREEGTWSTIIDGQ---KQHGSLSDKROYETADMRKBDTAESIDLNQKSTAD 129
 258 AIOPLPEALQORHLYFLRYEDLVRAPIAQT-RLYK-----FVGLDFLPHLQTVYN 307
 DB RLINLDEIYEQ---CRQELVHTGMKAEQIYKAEKAVIDLAVGQDALHNSDAFLPH 185
 308 VT--RGKMGQHAFTNARNALNVSQAMRWSLPEYKVS---QLDAGCEAMDLLGYLYR 362
 DB PTLIDSGIGS-----SCLISDQTMVLPYSEFSSESLQKGC--TANILS-SSVR 234
 363 SQOEGNLSIDLSSSHLGOVFE 387
 DB 235 QKRELTYMTIEYFNSA---GQRVAE 256

RESULT 8

364K Golgi complex-associated protein - rat
 J05837
 C:Species: Rattus norvegicus (Norway rat)
 C>Date: 05-Mar-1998 #sequence_revision 13-Mar-1998 #text_change 20-Jun-2000
 C:Accession: J05837
 R:Toki, C.; Fujiwara, T.; Sohma, M.; Hong, H.S.; Misumi, Y.; Ikehara, Y.
 Cell Struct. Funct. 22, 565-577, 1997
 A:Title: Identification and characterization of rat 364-kDa Golgi-associated protein reo
 A:Reference number: J05837; MUID:98093490; PMID:9431462
 A:Accession: J05837
 A:Status: nucleic acid sequence not shown
 A:Molecule type: mRNA
 A:Residues: 1-3187 <TOK>
 A:Cross-references: DDBJ:D25343; NID:9516825; PIDN:BA05026.1; PID:9516826
 C:Comment: This protein plays a role in the formation and maintenance of the characteris
 C:Superfamily: giantin
 F:49-549,624-1176,1238-1707,1763-3114/Domain: coiled-coil leucine zipper #status predict
 F:3165-3187/Domain: membrane anchor #status predicted <MAD>

Query Match 4.8%; Score 98.5; DB 2; Length 3187;
 Best Local Similarity 19.9%; Pred. No. 14;
 Matches 55; Conservative 54; Mismatches 93; Indels 75; Gaps 12;

149 KHCKLGGQGFDPDW---EKACRSHGFVLEKREPLSL-----QALYPLTTPSLNLHV 200
 DB 2286 QNCKNCRQLETDLPASRELTLRLHDEINVEKQKITLSLGSKEEALIVAAIAE---LHQO 2341
 201 HL--VRDPAVFRSHHTIEL-----MVDSHVILGQHLETIKEEDQPYAMKIIICKSQ 252
 DB 2342 HSKELKELENLISQEEENLTLEENKRAVEKTNLTLETETKE-----SLEQKRAQ 2394
 253 VD-IYKAIOQLPE-----ALQORL--FLRYEDLVRAPIAQTRLYKFGVGLDFP 299
 DB 2395 LDFEYKSSSLQDDDDRIYSDYRQLEERHLSVILEKDELIDQAAENNNKKEI----- 2448
 300 HLGTFVYVNTGKGGQAHFHNARNA-----LNVSQAMRWSLPEYKVS 343

DB 2449 -----RGLRGHMDINSENAKIDAEILQYRRDLNVEYTIKDSQORLLEAOIQ 2496
 OY 344 QLQDAGCEAMDLLGYLYOVRSOQEOG-NISIDLLSSSH 379
 DB 2497 QNKELRNECVKLEGRKLGSEAKQSLQNSLDALQGEN 2533

RESULT 9

MC240 homolog F10.orf291 - Mycoplasma pneumoniae (strain ATCC 29342)
 S73826
 C:Species: Mycoplasma pneumoniae
 A:Variety: ATCC 29342
 C>Date: 27-Feb-1997 #sequence_revision 25-Apr-1997 #text_change 07-Dec-1999

C:Accession: S73826
 R:Himmelreich, R.; Hilbert, H.; Plagens, H.; Pirkl, E.; Li, B.C.; Herrmann, R.
 Nucleic Acids Res. 24, 4420-4449, 1996
 A:Title: Complete sequence analysis of the genome of the bacterium Mycoplasma pneumon
 A:Reference number: S73327; MUID:97105885; PMID:8948633
 A:Accession: S73826
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-291 <HIM>

A:Cross-references: EMBL:AE000049; GB:U00089; NID:91674188; PIDN:AB96148.1; PID:9167
 A>Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1996
 C:Genetics:
 A:Genetic code: SGC3
 A:Superfamily: hypothetical protein ygeK

Query Match 4.8%; Score 97.5; DB 2; Length 291;
 Best Local Similarity 22.8%; Pred. No. 0 72;
 Matches 59; Conservative 43; Mismatches 94; Indels 63; Gaps 14;

OY 109 DAYNMPGRKQSSLEPQW---EQSRALCSAPVDFE--PAHETSSPKHCKLGGQGFDPW 163
 DB 41 ELYPLIGSDKLSELAKMHNIEQLQKCRVCYERGRYPIDQLVQGFVRLFGKCPDLDA 100
 OY 164 -EKACRSHGFVLEKREPLSLQALYPLTTPSLNLHVHLYVLRPRAVFRSHHTIELMV 222
 DB 101 SSEMGSHNF---ROI-----PAKVLYHTH-----OH----- 124
 OY 223 DSHVILGQHLETIKEEDQPYAMKIIICKSQVDIVAIQTLPLAQRVFLRYEDLVRA 281
 DB 125 --NIIYKTLQTLDEPRAQCHLRV---GQLAKTLAVANKLDG-KTAYTAGVHDLAKOL 178
 OY 282 PLAQTRLYKFEVGLDFLPHLQTV--YVNTGKMGQHAFTNARNALNVSQAMRWSLPEY 340
 DB 179 POALEKELAKVAGVNDYP---SMKVLYHSYAGVYILKHWYGLN--NSAVFSALMNTVTPQ 233
 OY 341 KVSQLODAGCEAMDLLGYL 359
 DB 234 KMSQL-----DWIIVL 244

RESULT 10

F90866
 transcription regulator tyrr Ecs1902 [imported] - Escherichia coli (strain 0157:H7, s
 C:Species: Escherichia coli
 C>Date: 18-Jul-2001 #sequence_revision 18-Jul-2001 #text_change 03-Mar-2003
 C:Accession: F90866

R:Hayashi, T.; Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C
 gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shingawa, H.
 DNA Res. 8, 11-22, 2001
 A:Title: Complete genome sequence of enterohemorrhagic Escherichia coli 0157:H7 and 9
 A:Reference number: A99629; MUID:21156231; PMID:11258796
 A:Accession: F90866
 A:Status: preliminary

A:Molecule type: DNA
 A:Residues: 1-513 <HAY>
 A:Cross-references: GB:BA000007; PIDN:BA935325.1; PID:91361367; GSPDB:GN00154
 A:Experimental source: strain 0157:H7, substrain RIMD 0509952
 C:Genetics:

A:Gene: Ecs1902
 C:Superfamily: transcriptional regulator of Tyrr type with ACT, PAS, AAA, and FIS dom

RESULT 13

AD0362

probable lysR-family transcription regulatory protein YPO2979 [Imported] - Yersinia pestis
 C:Species: Yersinia pestis
 C>Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 02-Nov-2001

C/Accession: AD0362

R:Fairhill, J.; Wren, B.W.; Thomson, N.R.; Tibball, R.W.; Holden, M.T.G.; Prentice, M.B.; deno-Farraga, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.; 11, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrell, Nature 413, 523-527, 2001

A>Title: Genome sequence of Yersinia pestis, the causative agent of plague.

A:Reference number: AB0001; MID:21470413; PMID:11586360

A:Accession: AD0362

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-292 <CUR>

A:Cross-references: GB:AL590842; PIDN:GAC92223.1; PID:g15980935; GSPDB:GN00175

A:Gene: YPO2979

Query Match

Best Local Similarity 23.7%; Score 92.5; DB 2; Length 292; Pred. No. 2;

Matches 50; Conservative 25; Mismatches 77; Indels 59; Gaps 10;

QY 184 QALYPLTPSLNHYVHLVPRPRAVFRSREHT-----IE-----LWDSHYVG 229

DB 53 QARQPLVD-----HGRRLVAVOETLASERLTSLRIAGOVERLFLVSDYYQPA 107

QY 230 QHLETIKEEDOPYAMKIIIC--KSGVDYIKAIQTLPEA-----LQORFLFRYEDLVNA 281

DB 108 HHENLIRREQRYPRVEFCVLADEEDVDILQ--LKRATIGVEMQER-----PPDIAT 162

QY 282 PLAQTRLYKY-----GLDFPLQTVYVNTVRGKMGQHAFTNARNALNSQAWMSL 337

DB 163 RLAAQTEMAVFLRGHPLRLPHLQNEQATLR-----QLCLNTYRTRDHNSQGLTWSA 217

QY 338 PYEKVSQLODAGCEAMDILGYQVRSQDQSG 368

DB 218 P-----SYLMLEMAEQG 230

RESULT 14

transcription regulator tyrR - Escherichia coli (strain K-12)

C:Species: Escherichia coli

C>Date: 31-Mar-1988 #sequence_revision 05-Apr-1995 #text_change 03-Mar-2003

C/Accession: A47086; A24209; F64881

R:Yang, J.; Ganesan, S.; Sarsero, J.; Pittard, A.J.

J. Bacteriol. 175, 1767-1776, 1993

A>Title: A genetic analysis of various functions of the TyrR protein of Escherichia coli

A:Reference number: A47086; MID:93194802; PMID:8449883

A:Accession: A47086

A:Molecule type: DNA

A:Residues: 1-513 <YAN>

A:Cross-references: GB:M12114; NID:g148091; PIDN:AAA24706.1; PID:g148092

A>Note: sequence extracted from NCBI backbone (NCBIP:127622)

A:Note: correction of sequence reported in reference A24209

R:Corrishi, E.C.; Atgyropoulos, V.P.; Pittard, J.; Davidson, B.E.

J. Biol. Chem. 261, 403-410, 1986

A>Title: Structure of the Escherichia coli K12 regulatory gene tyrR. Nucleotide sequence

A:Reference number: A24209; MID:86085847; PMID:3001057

A:Accession: A24209

A:Molecule type: DNA

A:Residues: 1-452, 'AKMRKVRWTKSPAVLNARY' <COR>

A:Experimental source: strain K12

A>Note: this sequence has been corrected in reference A47086

R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.; C.

A.; Rose, D.J.; Mau, B.; Shao, Y.

Science 277, 1453-1462, 1997

A>Title: The complete genome sequence of Escherichia coli K-12.

A:Reference number: A4720; MID:97426617; PMID:9278503

A:Accession: F64881

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-513 <BIAT>

A:Cross-references: GB:AE000230; GB:U00096; NID:g1787578; PIDN:AAC74405.1; PID:g17875

A:Experimental source: strain K-12, substrain MG1655

A:Gene: tyrR

A:Map position: 29 min

A:Complex: homodimer

A:Description: regulates expression of genes involved in aromatic amino acid biosynth

A:Note: autogenously regulated; activity depends on concentrations of aromatic amino

C:Superfamily: transcriptional regulator of Tyr type with ACT, PAS, AAA, and FIS dom

C:Keywords: DNA binding; nucleotide binding; P-loop; transcription regulation

F:206-421/Domain: RNA polymerase sigma factor interaction domain homology <SF1>

F:234-241/Region: nucleotide-binding motif A (P-loop) #status atypical

F:294-298/Region: nucleotide-binding motif B

Query Match

Best Local Similarity 19.4%; Score 92.5; DB 1; Length 513; Pred. No. 4.2;

Matches 84; Conservative 71; Mismatches 157; Indels 121; Gaps 21;

QY 5 KKGRLMFGSQVIVVALEFIMSVHRHLSQRESRPVRVYVLTSSMRGSSFPVGLFGH 64

DB 116 KIDLRLNNTAOLINGFNL-----RWLESEPDSDHNEHYV-----NGQNLMEI----- 161

QY 65 PDVFLMEPAMVMTFTSTAMKLMMAVR--DLRSVFLCDMSVFDAYMNGPRKQSSL 122

DB 162 TPVYIQLDENDQHV-----LTGAVVMLRSTRMGRQLQNVAAQVSAFSGIVANSP--KMKHV 216

QY 123 FQWESRALCSAPVC-----DPF--PAHEISSPK-----HCKYLQCOQPFDMVEKAC 167

DB 217 VEQAKLMLASAPLLITGTGTGKDLFAVACHQ--ASPRGRKYLAIVNCAIPEDAVESEL 275

QY 168 RSH-----GTVLKEVRFSLQALYPLDPSLNLHVLVPRPRAVFRSR 213

DB 276 FGHAEGRKGFPEQANGSVLDELDEGMS-----PRQAKILFLND--GTFR-- 321

QY 214 EHTTLEMDSHVIGHLETIKEEDOPYAMKIIICKSGVDYIKAIQTLPEALQORYFL 273

DB 322 -----RVGEDHEVAVDVRVICAQKNLVELVQ--KMFREDELYX 358

QY 274 RYEDLV--RAPLAQTRLYKYFGLDPLQLQTVYVNTVRGKMGQHAFTNARNALNSQ 331

DB 359 RLNVLTLLNPLRLDCPQ-----DIMPTELEVARFADQGVPRPKLAADLNTVLT--- 408

QY 332 AMRWSLPEYKVSQLODAGCEAMDILGYQVRSQ-----QEOGNSLIDLIS 376

DB 409 --RYAMP--GNVQRLNATYRALTOLDGYELRQDILLDYDAATYAVGEDAMGESSLDELIT 465

QY 377 S---SHILGQVFR 386

DB 466 SRFERSVLTOLYR 478

RESULT 15

S61023

hypothetical protein YPL242c - yeast (Saccharomyces cerevisiae)

N:Alternate names: hypothetical protein P1041

C:Species: Saccharomyces cerevisiae

C>Date: 15-Feb-1996 #sequence_revision 01-Mar-1996 #text_change 19-Apr-2002

C/Accession: S61023; S65271

R:PoHL, T.M.

submitted to the EMBL Data Library, November 1995

A:Reference number: S61010

A:Accession: S61023

A:Molecule type: DNA

A:Residues: 1-1495 <POHL>

A:Cross-references: EMBL:Z67751; NID:g1061234; PID:g1061248

R:PoHL, T.M.

submitted to the Protein Sequence Database, May 1996

A:Reference number: S64899

A:Accession: S65271
 A:Molecule type: DNA
 A:Residues: 1-1495 <POW>
 A:Cross-references: EMBL:Z73598; NID:91370498; PID:e247077; PID:91370499; MIPS:YPL242c
 A:Experimental source: strain S286c (AB972)
 C:Genetics:
 A:Gene: SGD:IOG1
 A:Cross-references: SGD:S0006163
 A:Map position: 16L

Query Match 4.5%; Score 92.5; DB 2; Length 1495;
 Best Local Similarity 20.3%; Pred. No. 17;
 Matches 63; Conservative 42; Mismatches 81; Indels 125; Gaps 18;

```

QY 67 VEVYMEP---AMHWMTFTSTAKLMAVADLLRSV---FLQMSYFDAMNPGRRQS 120
   ||||| :||:||||| :||| :||| :||| :|||
Db 846 LVIYKEPEFYAKNVTMTFGT-----VQRMNDREKSYTFRFCEM--LQNAINEAPSIES 898
   ||||| :||:||||| :||| :||| :||| :|||
QY 121 SL-----FQWQSRLCS--APYCDFE-----PAHEISSPK 149
   ||||| :||:||||| :||| :||| :||| :|||
Db 899 FLQNRQGFQWTLLQDFLRRESPEFSIIVPLDYLSDPVDFESDPYKIQEIHGFSSPQ 958
   ||||| :||:||||| :||| :||| :||| :|||
QY 150 HCKLLCGQOPFDWVEKACRSHGFVYLKEVRFSLQALYPLTDPSLMHVHVHLVDPRAV 209
   ||||| :||:||||| :||| :||| :||| :|||
Db 959 HC-----SPVDASTKNKFTD-----NLRCIM-----HAIEWVAE---I 989
   ||||| :||:||||| :||| :||| :||| :|||
QY 210 FRSREHT-TIELWVDSHIVLGQHELIKEDQPYAMKIICKS-----QVDIVKATQT- 261
   ||||| :||:||||| :||| :||| :||| :|||
Db 990 YTRKVVHTIPVEIR-----YLCRTKIFCYADKNIEETDSLRAISSI 1029
   ||||| :||:||||| :||| :||| :||| :|||
QY 262 LPEALQQRVYFLR-----YE-----DLVRAPLAQFTRLYKFVGL-----DF 297
   ||||| :||:||||| :||| :||| :||| :|||
Db 1030 LVNVFVSEIYLVNREYGYGKDSNVQKNQKIDILMKSLATVFEIKNFDFGLDPLNQYANEI 1089
   ||||| :||:||||| :||| :||| :||| :|||
QY 298 LPHLQTVVYVY 308
   ||||| :||:||||| :||| :||| :||| :|||
Db 1090 KPHIKDVLVYV 1100
   ||||| :||:||||| :||| :||| :||| :|||
  
```

Search completed: August 9, 2003, 16:44:59
 Job time : 17.0388 secs

THIS PAGE BLANK (USPTO)

Matches 118: Conservative 67: Mismatches 133: Indels 43: Gaps 10:

QY 37 ESRRPVLVYSSWSSGSEFGQLFGPHDVEYLMPEAHVMMFTT---SSTAWKLHMA 92
 DB 110 EPRR--HVLMMATRTGSSVEGFEFNQGNIFELPELPHIRVTYFEPGAGNANVGSALV 167
 QY 93 VADLLRSVFLCDMSFADAMNPGPRKQ--SSLFQWDSRSLCSAVYCDFFPA-HEISSPK 149
 DB 168 YRDVLAQQLLDCDLYLLEFISIPAPDEHLLTAALFRGSSSHLCEEPVCT--PSLKVFEKY 225
 QY 150 HCK-FLCGQGFPMDEKACRSHGFVVLKEVRFSLQALYPLTDDSLNHLVHLVDRPA 208
 DB 226 HCKNRCCGPIINTLLAEACRKHMAKTVRIQLEPLQPLAEDRLDIRITQVLRDRA 285
 QY 209 VFRSEHTTIELMDSHIVLGOHLETFKEEDPYAMKICKSQVDYKAIQ----- 260
 DB 286 VLVSR-----MAFSGKYESMKK-----MAEGEAPLQDEVOQLNGCESIRL 329
 QY 261 -----TLPEALQORYLFLRYEDLVRAPIAOTTRLYKFGVDFLPHLOTVVYNTRGKM 314
 DB 330 SAEGLRQPRWIRGRTMLVREDVARAPLRKALEMTRRAGIHPTQVEEMIRANTQAP-Q 388
 QY 315 GQHAHTARNALNYSQAMRMSLPYKVSQLODAGCEAMDILGYLQVRSQOQNLSDL 374
 DB 389 DNGIYSTQKNSSSEQFEKWRFSIPKRLAQVQDACEPMRLGYKLASAGELTRISL 448
 QY 375 L 375
 DB 449 L 449

RESULT 2
 YM67_MYCTU STANDARD; PRT; 388 AA.
 AC Q50695;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Hypothetical protein RV2267C.
 GN RV2267C OR MT2329 OR MTCY339.43.
 OS Mycobacterium tuberculosis.
 OC Bacteria; Actinobacteria; Actinobacteriales; Actinomycetales;
 CC Corynebacteriaceae; Mycobacteriaceae; Mycobacterium.
 OX NCBI_TaxID=1773;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-H37RV;
 RX MEDLINE=98295987; PubMed=9634230;
 RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
 RA Gordon S.V., Eigmeier K., Gas S., Barry C.E. III, Tekala F.,
 RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
 RA Davies R., Devlin K., Feltwell T., Gentles S., Hamlin N., Holroyd S.,
 RA Hornsby T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
 RA Oliver S., Osborne J., Quail M.A., Rajandream M.A., Rogers J.,
 RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
 RA Sulston J.E., Taylor K., Whitehead S., Barrett B.G.;
 RT "Deciphering the biology of Mycobacterium tuberculosis from the
 RT complete genome sequence."
 RL Nature 393:537-544 (1998).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CDC 1551 / Oshkosh;
 RA Fleischmann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
 RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
 RA Kolonay J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
 RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
 RA Bishai W.;
 RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
 RT laboratory strains";
 RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-

CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sdb.ch/announce/>
 CC or send an email to license@sdb.ch).
 CC -----
 DR EMBL: 277163; CAB00968.1; -
 DR EMBL: AE007076; AAK4611.1; -
 DR PIR: G70729; G70729.
 DR TIGR: MT2329; -
 DR Tuberculist: RV2267C; -
 DR InterPro: IPR00863; Sulfotransferase.
 DR Pfam: PF00685; Sulfotransferase; 1.
 DR Hypothetical protein; Complete proteome.
 KW SEQUENCE 388 AA; 46062 MW; 5DEB0263275A9B24 CRC64;
 Query Match 6.6%; Score 134.5; DB 1; Length 388;
 Best Local Similarity 21.8%; Pred. No. 0.00032;
 Matches 77: Conservative 59: Mismatches 116: Indels 101: Gaps 20:

QY 44 VLVSSWSSGSEFGQLF---GQH-----PDVEYLMPEAHV--WMFTSTAWKL 89
 DB 83 IFIVGHWRTGTLTLLHLLVVDNRHTGPGYECIALHHPLE--WFAYVEFLVSK---- 136
 QY 90 HMAVDDLRSVFLCDMSFADAMNPGPRKQSSLPQWDSRSLCSAPVCDFFPAHSSPK 149
 DB 137 HRAQDNM-----DLSLH-----HPQDEDFW-CMGLSPYLTLTAEPNR--PQ 177
 QY 150 HCKLLCGQGFPMDEKACRSHGFVVLKEVRFSLQALY-----PLTDDPSLNLV---- 199
 DB 178 Y-----EEYLDLEQVAPRE--LEIMKRLFLFFQGVVFRKKVYILKNPTSPRIKVL 229
 QY 200 -----VHLVDRPAVFRSEHTTIELMDSHIVLGOHLETFKEEDPYAMKICKS 251
 DB 230 EYFPAKFIHIVRDPYVYPS---TIHLKALYRIHGLQPTFGGLDK-----VYST 279
 QY 250 QVDYKAIQTLPEALQ-QRYLFLRYEDLVRAPIAOTTRLYKVGVD-----FLPHLOTVWY 306
 DB 280 YVDLYRKLDEGRELVDPTRFELRYEDLGDGGLRRYQHLGLADECTYPRIRQY-- 337
 QY 307 NTVRGKMGQHA-FHTNARNALNYSQAMRMSLPYKVSQLODAGCEAMDILGY 358
 DB 338 -----LADHADYKTN-----YQLTVGRALVDBMKELIIDRQY 372

RESULT 3
 PKSM_BACSU STANDARD; PRT; 4273 AA.
 AC P40872; O31781;
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 28-FEB-2003 (Rel. 41, Last annotation update)
 DE Putative polyketide synthase pksM.
 GN PKSM OR PKSY
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacilliales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessières P., Bolotin A., Borchert S.,
 RA Borriss R., Boursier L., Brans A., Braun M., Brignelli S.C., Bron S.,
 RA Brouillet S., Brunsch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabelt C., Ferrati E., Foulger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghm S.Y., Glaser P., Goffeau A., Gollightly E.J., Grandi G.,
 RA Guiseppi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holsappel S., Hosono S., Hulo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,

Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M., Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V., Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C., Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M., Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H., Parro V., Pohl T.M., Portelle D., Porwollik S., Prescott A.M., Presecan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S., Rieger M., Rivolta C., Roehle E., Roche B., Rose M., Sadate Y., Sato T., Scanlan E., Schleich S., Schoeter P., Scofield F., Sekiguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B., Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K., Takeuchi M., Tamakoshi A., Tanaka T., Terpstra P., Tognoni A., Tosoato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A., Viari A., Wamput R., Wedler E., Wedler H., Wetzenecker T., Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K., Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Zanchin A., "the complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*," Nature 390:249-256(1997).

RT Nature 390:249-256(1997).

RL [2]

RN SEQUENCE OF 1-1763 FROM N.A.

RP STRAIN-168 / PH1424;

RA Tognoni A., Grand G.;

RL Submitted (JUL-1994) to the EMBL/GenBank/DBJ databases.

CC -1- FUNCTION: POTENTIALLY INVOLVED IN SOME INTERMEDIATE STEPS FOR THE SYNTHESIS OF A POLYKETIDE MOLECULE WHICH MAY BE INVOLVED IN SECONDARY METABOLISM.

CC -1- COFACTOR: CONTAINS 4 COVALENTLY BOUND PHOSPHOPANTETHEINES (POTENTIAL).

CC -1- SIMILARITY: Contains 4 acyl carrier domains.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC -----

DR EMBL: Z59113; CAB13603.1; -

DR EMBL: Z51133; CAB14505.1; -

DR PIR: C69679; C69679.

DR Subtilast; BG10931; pksM.

DR InterPro: IPR002198; ADH_short.

DR InterPro: IPR000794; Ketoacyl-synt.

DR InterPro: IPR001601; Methyltransf.

DR InterPro: IPR006163; pp_bind.

DR InterPro: IPR006162; pantine_attach.

DR InterPro: IPR000051; SAM_bind.

DR Pfam: PF00106; adh_short; 1.

DR Pfam: PF00109; ketoacyl-synt; 3.

DR Pfam: PF02801; ketoacyl-synt_C; 3.

DR Pfam: PF00550; pp-binding; 4.

DR PROSITE: PS00012; PHOSPHOPANTETHEINE; 2.

DR PROSITE: PS00606; B_KETOACYL-SYNTHASE; 2.

DR PROSITE: PS00075; ACP_DOMAIN; 4.

KW Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.

KM Phosphopantetheine; Multifunctional enzyme; Repeat; Complete proteome.

FT DOMAIN 295 364

FT DOMAIN 396 834

FT DOMAIN 2190 2258

FT DOMAIN 2322 2737

FT DOMAIN 3532 3947

FT DOMAIN 3410 3483

FT DOMAIN 4140 4209

FT BINDING 337 337

FT BINDING 2222 2222

FT BINDING 2476 2476

FT BINDING 3446 3446

FT ACT_SITE 3690 3690

FT BINDING 4172 4172

FT BINDING 103 103

FT CONFLICT 276 276

Q -> E (IN REF. 2).

FT CONFLICT 289 289 T -> S (IN REF. 2).

SO SEQUENCE 4273 AA; 477459 MM; 3BBFCFLA250AEB5A CRC64;

Query Match 5.0%; Score 101.5; DB 1; Length 4273;

Best Local Similarity 22.3%; P-adj. No. 5.3;

Matches 59; Conservative 44; Mismatches 99; Indels 63; Gaps 14;

QY 153 LGGGQPFPMVNERACSHGFVVLKEVRFSLALYPLNDPS---LNLVHVHLVDPRA 208

DB 25 LPLGLAYIDLIVQVDEHGT-AYQELKNTLITFYLIDESTIALTHV----- 74

QY 209 VRSRHTTIELVDSHIVLQGLTETKE-----EDQPYVAMKIKSQYDIK 257

DB 75 --SEREGWSTIIIDQ---KQGESLSKQRYETADMRKQETAFASIDLNQWSTAD 129

QY 258 AIGTLEPAAQRLFRYEDLYRAPLAQTT-RYK-----FGLDFPLHQTWYN 307

DB 130 RLNLDEIYEQ---CRSQELVHTGMKAKGQYEKAGAVIDLAVGOALRHSDFRH 185

QY 308 VT--RGKGQGHAFHTNARNALNVSAWRMSLPYEKVS--OLQDAGCAMDLGLQYR 362

DB 186 PTLIDSGSIGS-----SCLISDQTMPLPLYESFSASERLQKC--TARILS-SSVR 234

QY 363 SQEEDGNLSIDLSSSHILGOVFE 387

DB 235 OKKELTYMTIEYFNSA---GQKVAE 256

RESULT 4

ID Y240_MYCPN STANDARD; PRT; 344 AA.

AC P75442;

DT 01-NOV-1997 (Rel. 35, Created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 28-FEB-2003 (Rel. 41, Last annotation update)

DE Hypothetical protein MG240 homolog (F10_Ori291).

GN MFN336 OR MF500.

OS Mycoplasma pneumoniae.

OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.

OX NCBI_TaxID=2104;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN-ATCC 29342 / M129;

RX MEDLINE=97105885; PubMed=8948633;

RA Himmelfarb R., Hilbert H., Plagens H., Pirkl E., Li B.-C., Hermann R.;

RT "Complete sequence analysis of the genome of the bacterium *Mycoplasma pneumoniae*," Nucleic Acids Res. 24:4420-4449(1996).

RL Nucleic Acids Res. 24:4420-4449(1996).

RN [2]

RP REVISIONS.

RX MEDLINE=20411492; PubMed=10954595;

RA Dandekar T., Huynh M., Regula J.T., Ueberle B., Zimmermann C.U., Andrade M.A., Doerks T., Sanchez-Pulido L., Snel B., Suyama M., Yuan Y.P., Herrmann R., Bork P.;

RT "Re-annotating the *Mycoplasma pneumoniae* genome sequence: adding RT value, function and reading frames," Nucleic Acids Res. 28:3278-3288(2000).

RL Nucleic Acids Res. 28:3278-3288(2000).

CC -1- SIMILARITY: TO B.SUBTILIS YOEK.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).

CC EMBL: AE000049; AAB96148.2; -

CC HAMAP: MF_00244; atypical/fused; 1.

DR InterPro: IPR005249; Cons_hypoth488.

DR InterPro: IPR00674; HD.

DR InterPro: IPR003607; Met_phosphohydro.

-1- SIMILARITY: Contains 1 sigma-54 factor interaction ATP-binding domain.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC
 CC EMBL: M12114; AAA24706.1; -
 CC EMBL: AE000230; AAC74405.1; -
 CC EMBL: D90770; BAAL1905.1; -
 CC EMBL: D90771; BAAL1916.1; -
 CC PIR: A47086; RGECAV.
 CC Ecogene: EG11042; tyrR.
 CC InterPro: IPR003593; AAA_ATPase.
 CC InterPro: IPR002912; ACT.
 CC InterPro: IPR002197; HTH_Fls.
 CC InterPro: IPR000014; PAS_domain.
 CC InterPro: IPR002078; Sig54_Interact.
 CC Pfam: PF01842; ACT.1
 CC Pfam: PF00158; Sigma54_activat; 1.
 CC SMART: SM00382; AAA; 1.
 CC SMART: SM00091; PAS; 1.
 CC TIGRfam: TIGR01199; HTH_fls; 1.
 CC PROSITE: PS0112; PAS; 1.
 CC PROSITE: PS00675; SIGMA54_INTERACT_1; 1.
 CC PROSITE: PS00676; SIGMA54_INTERACT_2; 1.
 CC PROSITE: PS00688; SIGMA54_INTERACT_3; 1.
 CC PROSITE: PS00045; SIGMA54_INTERACT_4; 1.
 CC KEGG: Transcription regulation; Aromatic hydrocarbons catabolism;
 CC DNA-binding; Repressor; Activator; ATP-binding; Complete proteome.
 CC FT DOMAIN 78 149 PAS.
 CC FT NP_BIND 206 428 SIGMA-54 FACTOR INTERACTION (POTENTIAL).
 CC FT NP_BIND 234 241 ATP (POTENTIAL).
 CC FT NP_BIND 290 299 ATP (POTENTIAL).
 CC FT DNA_BIND 483 502 H-T-H MOTIF (POTENTIAL).
 CC SQ SEQUENCE 513 AA; 57656 MW; B580A401CA27866 CRC64;

Query Match 4.58; Score 92.5; DB 1; Length 513;

Best Local Similarity 19.48; Pred. No. 2.1; Matches 84; Conservative 71; Mismatches 157; Indels 121; Gaps 21;

CC 5 KKRILMFLGSOYIVVALFTIMSVYRHLSQRESRVRVHLVLSMRSSSPFGOLGQH 64
 CC 116 KDRLRKHTAQAOLINGNFL-----RWLESEPODSHNEHYI-----NGQFLEI----- 161
 CC 65 PDVFLMEPAMHVMFTSTSTANKLHMAVR--DLRSVFLCDMSVFDAYMNGPRKOSL 122
 CC 162 TPVYLDDENDQHV---LTGAVVMLRSTIMRGQLGVAAQDVSAFQIYAVSP--KKKHV 216
 CC 123 FQWQDSALCSAYVC-----DFF--PAHEISSPK---HCKLLCGQDPEDVEKAC 167
 CC 217 VEQAQKLAISAPLLITGDTGTGKDLFAVACHQ--ASPRACKPYLALNCASIPEDAVSEL 275
 CC 168 RSH-----GFVLEKVRFLSQALYPLLTDSLNHHVHLVRDRAVRSR 213
 CC 276 FGAPREGKGFEEQANGSVLLDEIGEMS-----PRMAKTLRFIND--GTR-- 321
 CC 214 EHTTIELMDSHIVLGQHELTIREDDPYAMKIKCSQYDIYAKIOTLPEALQORFL 273
 CC 322 -----RVEDGHEVHDVAVICATQKNVELVQ---KGMRELLY 358
 CC 274 RYEDLV--RAPLAQTRLYFVGLDPLHQTWVYVNTRGKGGQAHFTNARNALVNSQ 331
 CC 359 RLNVLTNLNPLRDCQ-----DIMPTELFVARFADQGVPRPLADLMTVLT--- 408
 CC 332 AMWMSLPYEKVSOLODACGEAMDLLGLQYRSQ-----QCGNLSDDL 376
 CC 409 --RYAMP--GNVRLKNAITRALTLQLODGYELRPDILLPDYDAATVAVGEMSGSDEIT 465

CC 377 S---SHILGOVER 386
 CC 466 SRFERSVLTQLYR 478

RESULT 6
 CC MFD_BUCAI STANDARD; PRT; 812 AA.
 CC ID AC P57381; 09F456;
 CC DT 16-OCT-2001 (Rel. 40, Created)
 CC DT 16-OCT-2001 (Rel. 40, Last sequence update)
 CC DE 16-OCT-2001 (Rel. 40, Last annotation update)
 CC GN MFP OR BU294.
 CC OS Buchnera aphidicola (subsp. Acyrthosiphon pisum) (Acyrthosiphon pisum
 CC symbiotic bacterium).
 CC CC Bacteria; Proteobacteria; Gammaproteobacteria; Enterobacteriales;
 CC OC Enterobacteriaceae; Buchnera.
 CC OX NCBI_TaxID=118099;
 CC RN [1]
 CC RP SEQUENCE FROM N.A.
 CC RC STRAIN=Tokyo 1998;
 CC RX MEDLINE=20445173; PubMed=10939077;
 CC RA Shigenobu S., Watanabe H., Hattori M., Sakaki Y., Ishikawa H.;
 CC RT "Genome sequence of the endocellular bacterial symbiont of aphids
 CC Buchnera sp. Aps.";
 CC RL Nature 407:81-86(2000).
 CC CC -1- FUNCTION: NECESSARY FOR STRAND-SPECIFIC REPAIR. A LESION IN THE
 CC TEMPLATE STRAND BLOCKS THE RNA POLYMERASE COMPLEX (RNAP) (BY
 CC SIMILARITY).
 CC CC -1- SIMILARITY: IN THE N-TERMINAL SECTION; BELONGS TO THE UVRB FAMILY.
 CC CC -1- SIMILARITY: IN THE C-TERMINAL SECTION; BELONGS TO THE HELICASE
 CC CC FAMILY. RECG SUBFAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

CC EMBL: AP001118; BAB13004.1; -
 CC DR InterPro: IPR001410; DEAD.
 CC DR InterPro: IPR001650; Helicase_C.
 CC DR InterPro: IPR005118; TRCF.
 CC DR Pfam: PF02559; Card_TRCF; 1.
 CC DR Pfam: PF00270; DEAD; 1.
 CC DR Pfam: PF00271; Helicase_C; 1.
 CC DR Pfam: PF03461; TRCF; 1.
 CC DR SMART: SM00487; DEXDC; 1.
 CC DR SMART: SM00490; HELIC_C; 1.
 CC KW Helicase; DNA repair; ATP-binding; DNA-binding; Complete proteome.
 CC FT DOMAIN 263 633 HELICASE; RECG-LIKE.
 CC FT NP_BIND 293 300 ATP (POTENTIAL).
 CC FT SITE 394 397 DEEP BOX.
 CC SQ SEQUENCE 812 AA; 93881 MW; BB5C9F54D9EE2C35 CRC64;

Query Match 4.58; Score 92; DB 1; Length 812;

Best Local Similarity 17.88; Pred. No. 4.1; Matches 63; Conservative 74; Mismatches 140; Indels 76; Gaps 16;

CC 41 PVHVLVSSMRSSGSPFGOLGQHPDVFLMEPAMHVMFTSTSTANKLHMAVADLLRSV 100
 CC 206 PLHRLGGDDMNKKEKHKISKTYVDH-----AAQLHTYAKRESKTG-----A 247
 CC 101 FLCDMSVFDAYMNGPRKOSLFEQWDSRALCSAPVDFPAHEISSP--KXCKLLCGQP 159
 CC 248 FKXNIKEDYLFCDNCSKRTS-----DQNEWM-----KYLKMSKRPIMDRILGCDVG 296
 CC 160 FDNVEKACRSHGFVLEKVRFLSQALYPLLTDP-----SLNLAHV---HLVR 204
 CC 297 FGKTEIMARA--SFLAVSNKQVALIVPTTLAQGHYKFKIRFSNMPVNIINILSRFOTQ 355

```

OY 205 DPRAVERREHTIILANDSHIVLGCHLE-----TIKEEDQPYAMKTIK-----SOVD 254
DB 356 EDDLFKFKTKNRINIIITHTLLEKNIEMCSGLIIDEHREPVSKHEITIKKITSND 415
OY 255 IYKAIQT-LPEALQORYFLRYEDLVRAPLAQTIRLYKFEVGLDPLPHL--QGVYVNTNG 311
DB 416 ILLTATPTIPRTLNAMATGIKDLSTIAKPPAQRALIKTFI--OEYSPIILRKITLREISRG 474
OY 312 KGMGO-HAFTNARNALNVSQAMRSLPEKYSQDADGCEMDLIGYIQRVS 363
DB 475 ---GGVYIYKVKVONIMINIAERLSTILIP-----EASIKIGHQOMN 512

RESULT 7
HEMZ_SYNEL
ID 15-SEP-2003 (Rel. 42, Created)
OBDG6: 15-SEP-2003 (Rel. 42, Last sequence update)
DE Ferrochelatase (EC 4.99.1.1) (Prothome ferro-lyase) (Heme
DE synthetase).
GN HEMH OR TIR2216.
OS Synecchococcus elongatus (Thermosynechococcus elongatus).
OC Bacteria; Cyanobacteria; Chroococcales; Synecchococcus.
OX NCBI_Taxid=32046;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BP-1:
RX MEDLINE-2225144; PubMed-12240834;
RA Nakamura Y., Kaneko T., Sato S., Ikeuchi M., Katoh H., Sasamoto S.,
RA Watanabe A., Iriyuchi M., Kawashima K., Kimura T., Kishida Y.,
RA Kiyokawa C., Kohara M., Matsumoto M., Matsuno A., Nakazaki N.,
RA Shimo S., Sugimoto M., Takeuchi C., Yamada M., Tabata S.;
RA *Complete genome structure of the thermophilic cyanobacterium
RT Thermosynechococcus elongatus BP-1."
RL DNA Res. 9:123-130(2002).
CC -1- FUNCTION: Catalyzes the ferrous insertion into protoporphyrin IX.
CC -1- CATALYTIC ACTIVITY: Protoporphyrin + Fe(2+) -> prothome + 2 H(+).
CC -1- PATHWAY: Prothome biosynthesis; last step.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC -1- SIMILARITY: Belongs to the ferrochelatase family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL, AP005376; BAC09768.1; ALT_INIT.
DR HAMAP: MF_00323; ? 1.
DR Pfam: PF00762; Ferrochelatase; 1.
DR PROSITE: TIGR00109; hemh; 1.
DR PROSITE: PS00534; FERROCHELATASE; 1.
DR Porphyryin biosynthesis; Heme biosynthesis; Lyase; Iron;
KM Complete proteome.
FT METAL 197 IRON (BY SIMILARITY).
FT METAL 278 IRON (BY SIMILARITY).
SQ SEQUENCE 388 AA; 43876 MW; 31D82310E4339A9 CRC64;

Query Match 4.5%; Score 91.5; DB 1; Length 388;
Best Local Similarity 19.2%; Pred. No. 1.8;
Matches 68; Conservative 52; Mismatches 109; Indels 125; Gaps 17;

OY 52 SSSSF--VGQFGQHPD---VFYLMPEAMHWMTFTSSITAWKLHNAVRLDLSRVFLCDMS 106
DB 135 SSSSFRLLESILWNPDELOKIRYTLIPSWYHNPYVA-----AMADLIR-----Q 179
OY 107 VFDAVMNPGPKRSSLFQWEGSRALCSAPVCDFFPAHETISSPKHCKLLCGQOPFDWVEKA 166

```

```

DB 180 ELDRCPNPD-----EAVIFFSAHV--PKSYTEAG-DPYOEIIRA 217
OY 167 C-----RSHGVVLKEVRFSLQALYPLTTPSLNHYVHLVROBRAVRSREHT 217
DB 218 CVRLIMAAALNRNNAVLAQSRVGPVEWLPYTEVILEL-----AAQVYT 264
OY 218 IELANDSHIVLGCHLETTIKEEDQPYAMKTIKSOVDIYKAIQTPE-----ALQORYL 271
DB 265 LVVYVISTFV--SEHETIQEIDIER--REIAAEAGIEVRRPALNDHGFISALAQ----- 317
OY 272 FLRYEDLVRAPIAQTIRLYKFEVGLDPLPHLQGVYVNTNGKMGHAFHTNARNALNVSQ 331
DB 318 -----LVKEALAAPRTFAEVN-----OSRRKRYLYPQE 346
OY 332 AMRSLPEKYSQDADGCEAMDILGYIQRVSQSGNLSPLSSS---HILG 382
DB 347 RWEKMG-----TSAERMRNGRLAMIGFLA-----LMTLISGCGPLHMLG 386

RESULT 8
K406_HUMAN
ID K406_HUMAN STANDARD: PRT: 1089 AA.
AC Q43156; Q96A38; Q9BR47; Q9HAK0;
DT 28-FEB-2003 (Rel. 41, Created)
DT 28-FEB-2003 (Rel. 41, Last sequence update)
DE 15-SEP-2003 (Rel. 42, Last annotation update)
DE Protein KIA0406.
GN KIA0406.
OS Homo sapiens (Human)
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
OX NCBI_Taxid=9606;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE-brain;
RX MEDLINE-98116655; PubMed-9455477;
RA Ishikawa K.-I., Nagase T., Nakajima D., Seki N., Ohira M.,
RA Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT *Prediction of the coding sequences of unidentified human genes. VIII.
RT 78 new cDNA clones from brain which code for large proteins in
RT vitro."
RL DNA Res. 4:307-313(1997).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE-21638749; PubMed-11780052;
RA Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,
RA Jones M., Stavrides G., Almeida J.P., Babbage A.K., Baguley C.L.,
RA Bailey J., Barlow K.F., Bates K.N., Beard L.M., Beare D.M.,
RA Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,
RA Buck D., Burrill W.D., Butler A.P., Carder C., Carter N.P.,
RA Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,
RA Clegg S., Cobley V.E., Collier R.E., Connor R.E., Corry N.R.,
RA Coulson A., Coville G.J., Deadman R., Dhani P.D., Dunn M.,
RA Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,
RA Graham D.V., Griffiths C., Griffiths M.N.D., Gilliam R., Hall R.E.,
RA Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,
RA Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,
RA Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,
RA Levaesialho M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,
RA Marsh V.L., Martin S.L., McConnachie L.J., Mcley K., McMurry A.A.,
RA Milne S.A., Mistry A., Moore M.J.F., Mullikin J.C., Nickerson T.,
RA Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,
RA Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,
RA Rice C.M., Ross M.T., Scott C.E., Selva H.K., Showkeen R., Sims S.,
RA Skuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,
RA Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.R., Thorpe A.,
RA Tracey A., Tomanas A.C., Vaudin M., Wall M., Wallis J.M.,
RA Whitehead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,
RA Whiting L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,
RA Rogers J.;
RT *The DNA sequence and comparative analysis of human chromosome 20."
RL Nature 414:865-871(2001).

```

OY		256	AIOITLDEALQQRILFLRYEDYLVRPLAQTRTLKRYGVGLDFPHLQTWYVNTNRCKGNQH	317
DQ		318	AFHTNARNLNLSQAWRSLPYEKYSQLODACGEAMD---	LTAIGVRSQGEGNIS 371
DB		759	VLFALMALAAQ-----	-WPPDTGNGHLD-----QSAGEE 788
DO		789	GSHLRNRP-----	ALEKSTTTEMEDIEQLNLTKRKDVAD-GNVS 828
RESULT 9.				
PIDU_HUMAN				
ID	PIDU_HUMAN	STANDARD:	PRT:	435 AA.
AC	OGH490; OBNZEF;			
DT	15-SEP-2003 (Rel. 42, Created)			
DT	15-SEP-2003 (Rel. 42, Last sequence update)			
DT	15-SEP-2003 (Rel. 42, Last annotation update)			
DE	Cell division cycle protein 91-like 1 (CDC91-like 1 protein) (PIG-U).			
GN	CDC91L1.			
OS	Homo sapiens (human).			
OC	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
OX	NCBI_Taxid=9606;			
PE	[1]			
RC	SEQUENCE FROM N.A.			
RC	TISSEB-Embryo;			
RA	Hong Y., Kinoshita T.;			
RL	"Human PIG-U and yeast Cdc91p are fifth subunits of GPI			
RL	transamidase.";			
RN	submitted (JUN-2002) to the EMBL/genbank/DBDJ databases.			
RN	[2]			
RP	SEQUENCE FROM N.A.			
RX	MEDLINE-21638749; PubMed-11780052;			
RA	Deloukas P., Matthews L.H., Ashurst J., Burton J., Gilbert J.G.R.,			
RA	Jones M., Scavrides G., Almeida J.P., Babbage A.K., Baggaley C.L.,			
RA	Bailey J., Barlow K.F., Bates K.N., Bead L.M., Beare D.M.,			
RA	Beasley O.P., Bird C.P., Blakey S.E., Bridgeman A.M., Brown A.J.,			
RA	Buck D., Buttrill W.D., Butler A.P., Carder C., Carter N.P.,			
RA	Chapman J.C., Clamp M., Clark G., Clark L.N., Clark S.Y., Clee C.M.,			
RA	Clegg S., Cobley V.E., Collier R.E., Connor R.E., Cordy N.R.,			
RA	Coulson A., Coville G.J., Deaman R., Dhali P.D., Dunn M.,			
RA	Ellington A.G., Frankland J.A., Fraser A., French L., Garner P.,			
RA	Graham D.V., Griffiths C., Griffiths M.N.D., Gwilliam R., Hall R.E.,			
RA	Hammond S., Harley J.L., Heath P.D., Ho S., Holden J.L., Howden P.J.,			
RA	Huckle E., Hunt A.R., Hunt S.E., Jekosch K., Johnson C.M., Johnson D.,			
RA	Kay M.P., Kimberley A.M., King A., Knights A., Laird G.K., Lawlor S.,			
RA	Lehvesmaa M.H., Leversha M.A., Lloyd C., Lloyd D.M., Lovell J.D.,			
RA	Marsh V.L., Martin S.L., McConachie I.J., McLay K., McMurray A.A.,			
RA	Milne S.A., Mistry D., Moore M.J.F., Mullikin J.C., Nickerson T.,			
RA	Oliver K., Parker A., Patel R., Pearce T.A.V., Peck A.I.,			
RA	Phillimore B.J.C.T., Prathalingam S.R., Plumb R.W., Ramsay H.,			
RA	Rice C.M., Ross M.T., Scott C.E., Sehra H.K., Shownkeen R., Sims S.,			
RA	Stuce C.D., Smith M.L., Soderlund C., Steward C.A., Sulston J.E.,			
RA	Swann R.M., Symamore N., Taylor R., Tee L., Thomas D.W., Thorpe A.,			
RA	Tracey A., Tromans A.C., Vaudin M., Wall M., Wallis J.M., Williams S.A.,			
RA	Whithead S.L., Whitaker P., Willey D.L., Williams L., Williams S.A.,			
RA	Wilmink L., Wray P.W., Hubbard T., Durbin R.M., Bentley D.R., Beck S.,			
RA	Rogers J.;			
RT	"The DNA sequence and comparative analysis of human chromosome 20.";			
RT	Nature 414:865-871(2001).			
RN	[3]			
RP	SEQUENCE FROM N.A.			
RC	TISSEB-Embryo;			
RA	Ota T., Nishikawa T., Suzuki Y., Kawai-Hio Y., Hayashi K., Ishii S.,			
RA	Saito K., Yamamoto J., Wakamatsu A., Nagai T., Nakamura Y.,			
RA	Nagahari K., Sugano S., Isogai T.			
RL	Submitted (MAR-2002) to the EMBL/genbank/DBDJ databases.			
RA	[4]			
RP	SEQUENCE FROM N.A.			
RC	TISSEB-Brain;			
RX	MEDLINE-23288257; PubMed-12477932;			
RA	Strausberg R.L., Feingold E.A., Gronow L.H., Derge J.G.,			
RA	Altschuler R.D., Collins F.S., Wagner L., Schmeien C.M., Schuller G.D.,			
RA	Klausner S.F., Zeeberg B., Buecaw K.H., Scheiner C.F., Bhat N.K.,			

RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stopleon M., Soares M.B., Bonaldi M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carlini P., Prange C.,
 RA Rabe S.S., Locuelli N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Boak S.A., McKernan P.J., McKernan K.J., Malek J.A., Gamarale P.H.,
 RA Villalón D.K., Muzny K.C., Hale S., Garcia A.M., Gay L.J., Hulik S.W.,
 RA Fahley J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smalhus D.E.,
 RA Scheraga A., Schein J.E., Jones S.J.M., Marra M.A.,
 RT Generation and initial analysis of more than 15,000 full-length
 RL human and mouse cDNA sequences.";
 CC Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 CC -1- FUNCTION: May play a role in the biosynthesis of GPI anchors.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
 CC -1- SIMILARITY: Belongs to the CDC91 family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, AB06842; BACS1626.1; -
 DR EMBL, AL118520; CAC14080.1; -
 DR EMBL, AK075507; BAC11660.1; -
 DR EMBL, BC030512; AAH30512.1; -
 DR GeneW, HGNC:15791; CDC91L1.
 DR KW Transmembrane.
 FT TRANSMEM 1 21 POTENTIAL.
 FT TRANSMEM 66 86 POTENTIAL.
 FT TRANSMEM 166 186 POTENTIAL.
 FT TRANSMEM 188 208 POTENTIAL.
 FT TRANSMEM 237 257 POTENTIAL.
 FT TRANSMEM 259 279 POTENTIAL.
 FT TRANSMEM 286 306 POTENTIAL.
 FT TRANSMEM 313 333 POTENTIAL.
 FT TRANSMEM 355 375 POTENTIAL.
 FT TRANSMEM 386 406 POTENTIAL.
 FT CONFLICT 407 407 Y -> N (IN REF. 3).
 SQ SEQUENCE 435 AA; 50051 MW; 65FDD/96AB533027 CRC64;
 Query Match 4.3%; Score 87; DB 1; Length 435;
 Best Local Similarity 23.9%; Pred. No. 5;
 Matches 52; Conservative 31; Mismatches 73; Indels 62; Gaps 10;
 QY 9 LLMELGSOVIVVALLFIHMSVHRHLSQRESRPRPVHLVLSMR-----SGGSFV 57
 DB 5 LVLVAVVAVTVAALFRSLAEFISERVEVSP-----LSMKRYVEGLSLDLGVSPYS 59
 QY 58 GQLFQGHQDVY-----LMEPRHVMVMTFTSSAMKLHVAVRDLASVFLCDMSVDAVWN 113
 DB 60 GAVFETPLITYLHFLIDYALVEMTIDALTAIALYAIODFNKRVF----- 107
 QY 114 PGRFOSSILFQWQSRALCSAP-VCDFF--PAHEISSPKHCKLLCGQGFDMWEKACRSH 170
 DB 108 ---KKQKLLLELDQ-----YAPDAVELIRTPHEMKRYIPDKALAFYLLNPTYLCSVANKST 159
 QY 171 -----GFVVLKEVR-----FLSL-----QALYPL 189
 DB 160 CAINTLLAFILITIKGSAPLSAFLALATAYOSLYPL 197
 RESULT 10
 VAS1_MOUSE STANDARD; PRT; 463 AA.
 AC 09RIQ9;

DT 16-OCT-2001 (Rel. 40; Created)
 DT 16-OCT-2001 (Rel. 40; Last sequence update)
 DT 28-FEB-2003 (Rel. 41; Last annotation update)
 DE Vacuolar ATP synthase subunit S1 precursor (EC 3.6.3.14) (V-ATPase S1
 DE subunit) (V-ATPase S1 accessory protein) (V-ATPase Acs5 subunit) (C7-1
 DE protein).
 GN ATP6P1 OR ATP6S1.
 OS Mus musculus (Mouse).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Rodentia; Sclerognathii; Muridae; Murinae; Mus.
 OX NCBI_TaxID=10090;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Hayashi A., Hattori A., Okaze H., Kozuma S., Seki N., Saito T.,
 RL Submitted (Aug-1999) to the EMBL/GenBank/DBJ databases.
 CC -1- FUNCTION: VACUOLAR ATPASE IS RESPONSIBLE FOR ACIDIFYING A VARIETY
 CC OF INTRACELLULAR COMPARTMENTS IN EUKARYOTIC CELLS (BY SIMILARITY).
 CC -1- CATALYTIC ACTIVITY: ATP + H(2)O + H(+) (in) -> ADP + phosphate +
 CC H(+) (out).
 CC -1- SUBUNIT: COMPOSED OF AT LEAST 10 SUBUNITS.
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Vacuolar (By
 CC similarity).
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC -----
 DR EMBL, AB031290; BAB83498.1; -
 DR MGD; MG1:109629; At6p1.
 DR KW ATP synthetase; Hydrogen ion transport; Hydrolase; ATP-binding;
 KW Transmembrane; Glycoprotein; Signal.
 FT SIGNAL 1 32 POTENTIAL.
 FT CHAIN 33 463 VACUOLAR ATP SYNTHASE SUBUNIT S1.
 FT TRANSMEM 43 433 POTENTIAL.
 FT CARBOHYD 164 164 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 255 255 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 267 267 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 290 290 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 297 297 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 344 344 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 351 351 N-LINKED (GLCNAC...) (POTENTIAL).
 FT CARBOHYD 359 359 N-LINKED (GLCNAC...) (POTENTIAL).
 SQ SEQUENCE 463 AA; 51007 MW; AE28D99718BA0AC0 CRC64;
 Query Match 4.2%; Score 85.5; DB 1; Length 463;
 Best Local Similarity 23.4%; Pred. No. 7.4; Length 463;
 Matches 88; Conservative 33; Mismatches 112; Indels 143; Gaps 22;
 QY 89 LHMAYRDLRLSYFLCD-MSV--FDAYMNPGRKQSLFQWQSRALCSAPVCDFFPA-- 142
 DB 76 LELGRANVL--LELDKLSIEDFTAYGVGFKQDSAPS--NLENMLDLAPSLVPAVDW 132
 QY 143 HEIS-----SRHCKLLCGQGFDMWEKACRSHGVYLVKEVRLSLALYPL 190
 DB 133 YAISTLTLYLOEKLGASPLHYDL-----ATLKE--LKLNSLPL 170
 QY 191 TDPSLNHLVHLVRPRAVFRREHTTIELWDSHIVLGOLHETKEEDOP--YAMKTIIC 249
 DB 171 L-----LIRLPYTA--SSGLMAPREVLTGNDVIGVYLSLTKSEDPYTLALTVAR 219
 QY 250 KSOV--DIKVAIOTLPEALQ-----RYLF-----LRYED----- 277
 DB 220 PSRAVDITMWAGGLGRLQLOVAPSPAIHPVSYNDAPRLLEFAQNFSAVAKKEMKDL 279
 QY 278 -----LVRAPLAQTTLYKFF--GLDFLP--HLQTVV----- 305
 DB 280 TSLFGEVNLNLTSFMDNSFAMLSLTYEPLGATVTEFKILASRFYVPSARYPMARL 339
 QY 306 -----YVWTGKMGQHAHF-----TNRNAL--NVSQAMKMSLPYEVKSQLO 346

DB 340 EIHNSGSAHFNFVSOVTPSYFHCYVSSVSKGNLVTNPSVMQMLHNQI-QAF 398
 QY 347 DACGE-----AMDLG 358
 DB 399 NVTGEQFSVSDCAF 414

RESULT 11
 SIM4_YEAST
 ID SIM4_YEAST STANDARD: PRT: 974 AA.
 AC P32259;
 DT 01-OCT-1993 (Rel. 27, Created)
 DT 01-OCT-1993 (Rel. 27, Last sequence update)
 DT 01-OCT-1996 (Rel. 34, Last annotation update)
 DE Global transcriptional regulator SIM4.
 GN SIM4 OR TSF3 OR BEL2 OR GAL22 OR SSF5 OR YNL236W OR N1135.
 OS Saccharomyces cerevisiae (Baker's yeast).
 OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
 OC Saccharomycetales; Saccharomycetaceae; Saccharomyces.
 CX NCBI_TaxID=4932;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=YM256;
 RX MEDLINE=93140781; PubMed=8423805;
 RA Chen S., West R.W. Jr., Johnson S.L., Gans H., Kruger B., Ma J.;
 RT "TSF3, a global regulatory protein that silences transcription of
 RT yeast GAL genes, also mediates repression by alpha 2 repressor and is
 RT identical to SIM4.";
 RL M01. Cell. Biol. 13:831-840(1993).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=93024394; PubMed=1406639;
 RA Jiang Y.W., Stillman D.J.;
 RT "Involvement of the SIM4 global transcriptional regulator in the
 RT chromatin structure of Saccharomyces cerevisiae.";
 RL M01. Cell. Biol. 12:4503-4514(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Harashima S., Mabuchi H., Ramash R., Hasebe M., Tanaka A., Oshima Y.;
 RL Submitted (MUG-1992) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=97051596; PubMed=8896273;
 RA Pandolfo D., de Antoni A., Lanfranchi G., Valle G.;
 RT "The DNA sequence of cosmid 14-5 from chromosome XIV reveals 21 open
 RT reading frames including a novel gene encoding a globin-like
 RT domain.";
 RL Yeast 12:1071-1076(1996).
 CC -1- FUNCTION: GLOBAL REGULATORY PROTEIN THAT SILENCES TRANSCRIPTION
 CC OF GAL AND MATING-TYPE GENES. NEGATIVE REGULATOR OF THE HO
 CC (HOMOTHALISM) GENE. MAY POTENTIATE TRANSCRIPTIONAL ACTIVATION
 CC AND REPRESSION BY REGULATING THE ACTIVITY OF THE BASAL
 CC TRANSCRIPTIONAL APPARATUS. SIM4 MUTATION ALTERS CHROMATIN
 CC STRUCTURE.
 CC -1- SUBCELLULAR LOCATION: Nuclear.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).

DR EMBL: M93050; AAA35044.1;
 DR EMBL: X64516; CAA45819.1;
 DR EMBL: D12918; BAA02302.1;
 DR EMBL: Z69381; CAA93362.1;
 DR EMBL: Z71512; CAA96140.1;
 DR PIR: A44484; A44484.
 DR TRANSFAC: T01243;
 DR SGD: S0005180; SIM4.

DR GO: GO:0000119; C:mediator complex; IDA.
 DR GO: GO:0016455; F:RNA polymerase II transcription mediator ac. .; IDA.
 DR GO: GO:0006366; P:transcription from Pol II promoter; IDA.
 KW Transcription regulation; Activator; DNA-binding; Repressor;
 KW Nuclear protein.
 FT DOMAIN 63 81 SER/THR-RICH.
 FT DOMAIN 889 893 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
 SO SEQUENCE 974 AA; 111296 MW; 12ECF5EACDC05A8E CRC64;
 Query Match 4.28; Score 85.5; DB 1; Length 974;
 Best Local Similarity 17.7%; Pred. No. 19;
 Matches 65; Conservative 49; Mismatches 131; Indels 123; Gaps 15;

QY 24 IHMSVHRLSORESRPRPVHVLSSRRSSFPGL- -FGQH- - - - -POVEF-LMPPAW 75
 DB 51 LHPOLHEVOYOEBSSTLSTSTTSTVSGSTAGVSTPNNGSNKSPQPFYNISSHW 110
 QY 76 HWMTF-----TSSWAKLHNAVRDLRSVFLCDMSVDAY 111
 DB 111 NNPFSLPGDMLAVODELGNMTMLTQGRPRATTYKLTWVFD- - - - -NVYKIY 160
 QY 112 MNPEP-----KRSSL- - - - -FGW-EQSRALCSAPVCF- - - - - 139
 DB 161 NHVWPLRPVDKLPNMNERQTRKEYNTSILEFRLTSSKSVISQPCARDSSNTYRSR 220
 QY 140 - - -FPAHEISSPKHCKLLC- - - - -GOQPPDMWEKACRSHGFVL- - - - -KEVRF 180
 DB 221 AQGVPPYGVHPPIPTACLAIRKNGQIDFWQFSNKRDKKTKITQLDTSNORFQDLW 280
 QY 181 LSLQALYPLLTDPSSL- - - - -NLHV- - - - -VHLVDPRAVRSREHTT 217
 DB 281 LEFARITPMNDQCMILITTSYSKLSKNISFYKLVHNMVLMNTPRVLNDPSLKIQFILSTT 340
 QY 218 IELMVD-SHYLGCHLETI- - - - -KEEDPYAMKTIQCSQVDIVAIQTLEPAQRY 270
 DB 341 LDPTDGHVJLKEMLHNVSKSSTEKDPSEIILVYVCDTSKSLVRYRLAPQLSAEY 400
 QY 271 LFLRYEDL 278
 DB 401 LVILKPOL 408

RESULT 12
 COG7_DROME
 ID COG7_DROME STANDARD: PRT: 742 AA.
 AC OGYAD6; O85ZY9;
 DT 28-FEB-2003 (Rel. 41, Created)
 DT 28-FEB-2003 (Rel. 41, Last sequence update)
 DT 15-SEP-2003 (Rel. 42, Last annotation update)
 DE Putative conserved oligomeric Golgi complex component 7.
 GN CG31040/CG7880.
 OS Drosophila melanogaster (Fruit fly).
 OC Eukaryota; Metazoa; Arthropoda; Hexapoda; Insecta; Pterygota;
 OC Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
 OC Ephydroidea; Drosophilidae; Drosophila.
 CX NCBI_TaxID=7227;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Berkeley;
 RX MEDLINE=20196006; PubMed=10731132;
 RA Adams M.D., Celinker S.E., Holt R.A., Evans C.A., Gocayne J.D.,
 RA Amanatides P.G., Scherer S.E., Li P.W., Hoskins R.A., Galie R.F.,
 RA George R.A., Lewis S.E., Richards S., Ashburner M., Henderson S.N.,
 RA Sutton G.G., Wortman J.R., Yandell M.D., Zhang Q., Chen L.X.,
 RA Brandon R.C., Rogers Y.-H.C., Blazer R.G., Champe M., Pfeiffer B.D.,
 RA Wan K.H., Doyle C., Baxter E.G., Helt G., Nelson C.R., Miklos G.L.G.,
 RA April J.F., Agbayani A., An H.-J., Andrews-Pfannkoch C., Baldwin D.,
 RA Balow R.M., Basu A., Baxendale J., Bayraktaroglu L., Beasley E.M.,
 RA Beeson K.Y., Benos P.V., Bernan B.P., Bhandari D., Bolshakov S.,
 RA Borkov A., Botchan M.R., Bouck J., Brockstein P., Brotter P.,
 RA Burtis K.C., Busam D.A., Butler H., Cadieu E., Center A., Chandra I.,
 RA Cherry J.M., Cawley S., Dahlke C., Davenport L.B., Davies P.,
 RA de Pablos B., Delcher A., Deng Z., Mays A.D., Dew I., Dietz S.M.,

Query Match	4.2%	Score 85;	DB 1;	Length 742;
Best Local Similarity	22.1%	Pred. No. 15;		
Matches	53;	Conservative	40;	Mismatches 99; Indels 48; Gaps 8
QY	180	FLSLQALYPLLTDDSLNLHVHVLNR-----DPRAFRSREHTTT---ELMVDSHIVIG	229	
Db	288	FLVIAELLPAL-QPRDAHIILQLTKTSNERLEMLLFPKVNHSFVLIHLNLSLEQSHITTS	346	
QY	230	QHLTEIKEDDQPYAMKLT-----CKSOVDIVAIGLP-----EALQQRVL	271	
Db	347	ELHRLIGEAIFETFKFTIQYPRLEFQSLSTQVDRLSNQPTSPDGVARHLESTKLYE	406	
QY	272	FLAREDTLRAPLAQTTRLRYKEVGL---DELPHLQIVYVNVTRGKMGQGHAFHTNARNALN	328	
Db	407	WLKREACERCASITSDIALACKLITLLNGTIFKRQLESF-----GRIQQLDLSGSSSY	458	
QY	329	VSQAMRWLSLPYEKYSQLODACCGEAMDLLGYLQVRSQOEGNLSLDLLSSHTLGGVFRFG	388	
Db	459	AAOSEGNSLLQYTWSSQLO-----CLADFOVQLHGFEDQLHTRVMVLSNRLTRKPSNRG	510	

RESULT 13

APR2_YEAST STANDARD; PRT; 844 AA.

ID	ENTRY	STANDARD:	PRT:	844 AA.
AC	PEP2_YEAST			
AC	P32454;			
DT	01-OCT-1993 (Rel. 27, Created)			
DT	01-JUN-1994 (Rel. 29, Last sequence update)			
DT	28-FEB-2003 (Rel. 41, Last annotation update)			
DE	Aminopeptidase II (EC 3.4.11.-) (YSCII).			
GN	APPE2 OR LAP1 OR YKL157W OR YK1611.			
OS	Saccharomyces cerevisiae (Baker's yeast).			
OC	Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;			
OC	Saccharomycetales; Saccharomycetaceae; Saccharomyces.			
OX	NCBI_TaxID:4932;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-AB320;			
RC	MEDLINE-92111552; PubMed-1765107;			
RX	Garcia-Alvarez N., Cueva R., Suarez-Rendueles P.;			
RA	"Molecular cloning of soluble aminopeptidases from Saccharomyces			
RT	cerevisiae. Sequence analysis of aminopeptidase yscii, a putative			
RT	zinc-metallopeptidase.";			
RL	Eur. J. Biochem. 202:993-1002.(1991).			
RP	[2]			
RE	SEQUENCE FROM N.A.			
RC	STRAIN-S288C;			
RC	MEDLINE-94378720; PubMed-8091859;			
RA	Vandenbol M., Bolle P.-A., Dion C., Portetelle D., Hilger F.;			
RT	"DNA sequencing of a 36.2 kb fragment located between the FAS1 and			
RT	LAP1 loci of chromosome XI of Saccharomyces cerevisiae.";			
RL	Yeast 10:835-840.(1994).			
CC	-1- FUNCTION: IT IS INVOLVED IN THE CELLULAR SUPPLY OF LECITINE FROM			
CC	EXTERNALLY OFFERED LECITINE-CONTAINING DIPEPTIDE SUBSTRATES.			
CC	-1- COFACTOR: Binds 1 zinc ion (By similarity).			
CC	-1- SIMILARITY: Belongs to leucine peptidase family M1.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation			
CC	at the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sdb.ch/announce/ ,			
CC	or send an email to license@sdb-sib.ch).			
CC	-----			
DR	EMBL; X63998; CAA45403.1; -			
DR	EMBL; Z26877; CAA81497.1; -			
DR	EMBL; Z28157; CAA81999.1; -			
DR	PIR; S37794; S37794.			
DR	MEROPS; M01.006; -			
DR	SGD; S0001640; APE2.			
DR	GO; GO:0005737; C:cytoplasm; NAS.			
DR	GO; GO:0030287; C:periplasmic space (sensu Fungi); NAS.			

CC -1- ALTERNATIVE PRODUCTS:
CC Event-Alternative splicing; Named isoforms=2;
CC Name=1;
CC IsoId=Q8K358-1; Sequence-Displayed;
CC Note-No experimental confirmation available;
CC Name=2;
CC IsoId=Q8K358-2; Sequence-VSP_006955, VSP_006956;
CC Note-No experimental confirmation available;
CC -1- SIMILARITY: Belongs to the CDC91 family.
CC -1- CAUTION: The N-terminus is deduced from ESTs and the genomic
CC sequence.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL; AL845325; -; NOT ANNOTATED_CDS.
CC DR EMBL; BC013076; AAH13076.1; -;
CC DR EMBL; BC028278; AAH28278.1; ALT_INIT.
CC KW Transmembrane; Alternative splicing.
CC FT TRANSMEM 1 21 POTENTIAL.
CC FT TRANSMEM 66 86 POTENTIAL.
CC FT TRANSMEM 165 185 POTENTIAL.
CC FT TRANSMEM 187 207 POTENTIAL.
CC FT TRANSMEM 236 256 POTENTIAL.
CC FT TRANSMEM 258 278 POTENTIAL.
CC FT TRANSMEM 285 305 POTENTIAL.
CC FT TRANSMEM 312 332 POTENTIAL.
CC FT TRANSMEM 354 374 POTENTIAL.
CC FT TRANSMEM 385 405 POTENTIAL.
CC FT VARSPPLIC 350 357 FLNRITVL -> CEYSDFPH (in isoform 2).
CC FT VARSPPLIC 358 434 /FTId-VSP_006955.
CC FT VARSPPLIC 358 434 Missing (in isoform 2).
CC FT VARSPPLIC 358 434 /FTId-VSP_006956.
CC FT VARSPPLIC 358 434 Missing (in isoform 2).
CC SQ SEQUENCE 434 AA; 49804 MW; 901451A5A73C5515 CRC64;
CC
CC Query Match 4.18; Score 83.5; DB 1; Length 434;
CC Best Local Similarity 22.94; Pred. No. 10;
CC Matches 52; Conservative 36; Mismatches 60; Indels 79; Gaps 12;
CC
QY 17 VIVVAFIHMVHR-----HLSQRESRPVHVLVLSMR-----SGSFFYGOL 60
Db 8 VLVVAVTVRAALFRSSIAEFISERVEVSP-----LSSMKRYVEGLALLDLGVSPYSGAV 62
QY 61 FGGHPDVEY-----LMEPAHVMNTFTSTAWKLHMAVRDILRSVFLCDMSVFDAYMNGP 116
b 63 FHETPLIYFHFLLDYALVELVMTDALTAIALYFAIQDENKVF----- 107
QY 117 RKQSSLFQWESRALCSAPVCDFP--AHEISSPKCK-----LTCGQPFDMVEKACR 168
Db 108 KKQKLLLELDQ-----YAPDAVELIRTPMEMERYIPLKVALTLNPTILSCVAK 156
QY 169 SH-----GEVVLKEVR-----ELSL---QALXPL-LTDPSL 195
Db 157 STCAINNNTLAFILITTTIKGSVFLSAVFLATATYOSLYPVTLFAPGL 203

Search completed: August 9, 2003, 16:43:07
Job time : 11.5245 secs

STOP PAGE AND
REPLY TO THE BOARD
(USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 18:10:01 ; Search time 543 Seconds
(without alignments)
10156.449 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 2043
Sequence: 1 gaattccattggtgttggtta.....tgggatacttaaaaaaaaaa 2043

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 2552756 segs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 4193764

Minimum DB seq length: 15
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*

- 1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
- 2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
- 3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
- 4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
- 5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
- 6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
- 7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
- 8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
- 9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
- 10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
- 11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
- 12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
- 13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
- 14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
- 15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
- 16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
- 17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
- 18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
- 19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
- 20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
- 21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
- 22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
- 23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
- 24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
- 25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result NO.	Score	Query Match	Length	DB ID	Description
1	1547	75.7	1979	22	AAK94229 Human full-length
2	1547	75.7	2032	20	AAZ20792 Human glycosyl sul
3	1262	61.8	1333	24	AA516947 Human l-selectin s
4	1153	56.4	2065	21	AAZ94211 Human transferrase
5	695	34.0	877	22	AAK91803 Human cDNA 5'-end
6	695	34.0	877	22	AAK93921 Human cDNA clone t
7	403	19.7	505	24	ABV89280 Human colon cancer
8	403	19.7	517	24	ABK54724 Human colon cancer

9	306	15.0	2988	21	AAK76156 Human ORFX ORF1711
10	114	5.6	389	24	ABK54794 Human colon cancer
11	105	5.1	548	22	AAK92588 Human cDNA 3'-end
12	59	2.9	1647	24	AAK24670 Human drug metabol
13	59	2.9	1694	22	AAAD2700 Human glycosyl sul
14	59	2.9	2044	22	AAAD2699 Human glycosyl sul
15	59	2.9	2170	22	AAAD2698 Human glycosyl sul
16	59	2.9	2544	24	ABN69506 Human corneal N-ac
17	59	2.9	48436	24	ABN69533 Human corneal N-ac
18	59	2.9	160552	22	AAAD2697 Human glycosyl sul
19	37	1.8	37	20	AAZ20798 PCR primer for gly
20	30	1.5	30	24	ABK70821 Probe for human ge
21	27	1.3	1926	20	AAZ20793 Mouse glycosyl sul
22	23	1.1	1937	24	AA516948 Murine intestinal-
23	23	1.1	1989	22	AAAD2696 Mouse glycosyl sul
24	22	1.1	22	24	ABK70834 PCR primer for hum
25	22	1.1	22	24	AA516960 Human l-selectin s
26	22	1.1	22	24	AA516962 Human l-selectin s
27	22	1.1	22	24	AA516964 Human ovarian canc
28	22	1.1	167	24	ABL66480 Human ovarian canc
29	22	1.1	171	24	ABL66089 Human ovarian canc
30	22	1.1	187	24	ABL66207 Human ovarian canc
31	22	1.1	191	24	ABL66306 Human ovarian canc
32	21	1.0	21	24	ABK70835 PCR primer for hum
33	21	1.0	21	24	AA516961 Human l-selectin s
34	21	1.0	10004	22	ABK14483 Human nervous syst
35	21	1.0	10246	22	ABA14485 Human nervous syst
36	20	1.0	88	21	AAK25980 Human secreted pro
37	20	1.0	354	22	AAK65813 Novel human polynu
38	20	1.0	370	22	AAK65341 Novel human polynu
39	20	1.0	381	22	AAK67389 Bovine EST associa
40	20	1.0	398	25	ABK45934 Human myosin l-cha
41	20	1.0	3763	19	AAV58194 DNA encoding novel
42	20	1.0	3763	23	ABN69872 Human ovary specif
43	20	1.0	5814	24	ABN69855 Human dlthp protei
44	20	1.0	6147	25	ACC46453 Human immune/haema
45	20	1.0	9516	22	AAK73555

ALIGNMENTS

RESULT 1	AAK94229	standard; cDNA; 1979 BP.
ID	AAK94229	
AC	AAK94229;	
DT	06-NOV-2001	(first entry)
DE	Human full-length cDNA, SEQ ID NO: 2816.	
XX	Human; full length cDNA; cDNA synthesis; oligo-capping; ss.	
OS	Homo sapiens.	
PN	EP1130094-A2.	
XX	05-SEP-2001.	
PD	07-JUL-2000; 2000EP-0114089.	
XX	08-JUL-1999; 99JP-0194486.	
XX	PR 11-JAN-2000; 2000JP-0118774.	
PR	02-MAY-2000; 2000JP-0183765.	
XX	(HELI-) HELIX RES INST.	
PA	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;	
PI	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;	
XX	WPI: 2001-524255/58.	
DR	P-PSDB; AAM93309.	

XX 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Claim 8; SEQ ID NO 2816; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a full length
 CC human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 XX
 SQ Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other;

Query Match 75.78; Score 1547; DB 22; Length 1979;
 Best Local Similarity 99.98; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAAGGTCTTCACCTGACACAAATGCTACGCTTAAATAATGAAGCTCTGCTGTTCT 165
 DB 47 CAAGGTCTTCACCTGACACAAATGCTACGCTTAAATAATGAAGCTCTGCTGTTCT 106
 QY 166 GGTTCACAGATGAGCATCTTGGCTATCTTCCACATGTACAGCCACAACATCAGCTC 225
 DB 107 GGTTCACAGATGAGCATCTTGGCTATCTTCCACATGTACAGCCACAACATCAGCTC 166
 QY 226 CCGTCTATTAAGGACACAGCCGAGGCGATGACGCTGCTGCTTCTTCTGCGCTC 285
 DB 167 CCGTCTATTAAGGACACAGCCGAGGCGATGACGCTGCTGCTTCTTCTGCGCTC 226
 QY 286 TGCTCTCTCTTGTGGGAGAGCTTTTGGGAGAGCCAGATGTTTCTACCTGATGA 345
 DB 227 TGCTCTCTCTTGTGGGAGAGCTTTTGGGAGAGCCAGATGTTTCTACCTGATGA 286
 QY 346 GCCCGCTGACAGCTGTGATGACCTTCAAGCAGAGACCGCTGATGCTGACATGGC 405
 DB 287 GCCCGCTGACAGCTGTGATGACCTTCAAGCAGAGACCGCTGATGCTGACATGGC 346
 QY 406 TGTGGGAGATCTATACGGGCGCTTCTTGTGCGACATGAGCGCTTTGATGCTACAT 465
 DB 347 TGTGGGAGATCTATACGGGCGCTTCTTGTGCGACATGAGCGCTTTGATGCTACAT 406
 QY 466 GGAACCTGTCCTCCGAGAGACATCCAGCCCTTCAAGTGGAGAAACGCGGCGCTG 525
 DB 407 GGAACCTGTCCTCCGAGAGAGTCCAGCCCTTCAAGTGGAGAAACGCGGCGCTG 466
 QY 526 TTCTGACACCTGCTGTGACATCATCCACAAGATGAATCATCCCGGCGCTACATGAC 585
 DB 467 TTCTGACACCTGCTGTGACATCATCCACAAGATGAATCATCCCGGCGCTACATGAC 526
 QY 586 GGTCCGTGACAGTCAACAGCCCTTTAGAGTGTGGAGAAAGCGCTGCGCTCTACAGCA 645
 DB 527 GGTCCGTGACAGTCAACAGCCCTTTAGAGTGTGGAGAAAGCGCTGCGCTCTACAGCA 586
 QY 646 CGTGTGCTCAAGAGAGGTGGCTTCTTCAACCTGCAGTCCCTACCGCTGTGAAGA 705
 DB 587 CGTGTGCTCAAGAGAGGTGGCTTCTTCAACCTGCAGTCCCTACCGCTGTGAAGA 646
 QY 706 CCCCTCCCTCAACCTGCATATGCTGACACCTGTCCGGAGCCCGCGGCGCTGTCGTT 765
 DB 647 CCCCTCCCTCAACCTGCATATGCTGACACCTGTCCGGAGCCCGCGGCGCTGTCGTT 706
 QY 766 CCGAGAACGACAAAGAGAGATCTCATGATTGACAGTGCATTTGATGAGGAGCATGA 825
 DB 707 CCGAGAACGACAAAGAGAGATCTCATGATTGACAGTGCATTTGATGAGGAGCATGA 766

QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTACTATGTATGACGATCATCTGCCAAAGCA 885
 DB 767 GCAAAAACCTCAAGAGAGAGACCAACCTACTATGTATGACGATCATCTGCCAAAGCA 826
 QY 886 GCTGAGATCTACAAAGACCATCCAGTCTTGGCCCAAGGCGCTGAGAGAGCTACCTGCT 945
 DB 827 GCTGAGATCTACAAAGACCATCCAGTCTTGGCCCAAGGCGCTGAGAGAGCTACCTGCT 886
 QY 946 TGTGCGTATGAGAGACCTGGCTGAGGCGCTGAGGCGCCAGACTTCCGAATGTATGAAT 1005
 DB 887 TGTGCGTATGAGAGACCTGGCTGAGGCGCTGAGGCGCCAGACTTCCGAATGTATGAAT 946
 QY 1006 CGTGGGATTTGGAATTTTGGCCCATCTTTCAGACCTGGGTGATTAATCAACCCAGCA 1065
 DB 947 CGTGGGATTTGGAATTTTGGCCCATCTTTCAGACCTGGGTGATTAATCAACCCAGCA 1006
 QY 1066 GGGCATGGGAGACCAAGCTTTCACACAAATGGCAGGAGAGCCCTTAATGTCTCCAGGC 1125
 DB 1007 GGGCATGGGAGACCAAGCTTTCACACAAATGGCAGGAGAGCCCTTAATGTCTCCAGGC 1066
 QY 1126 TTGGCGCTGCTTGTGGCCATATGAAAGGTTTCTGACTTCAGAAACCTGTGGCGATGC 1185
 DB 1067 TTGGCGCTGCTTGTGGCCATATGAAAGGTTTCTGACTTCAGAAACCTGTGGCGATGC 1126
 QY 1186 CATGAATTTGCTGGGCTTACCGCCACGTCAGATCTGAAACAAAGAGAAACCTGTGCT 1245
 DB 1127 CATGAATTTGCTGGGCTTACCGCCACGTCAGATCTGAAACAAAGAGAAACCTGTGCT 1186
 QY 1246 GGATCTTCTGCTACCTGAGACTGCTGCTGAGCAATTCACATGAAGAGGTTGAGAAGCT 1305
 DB 1187 GGATCTTCTGCTACCTGAGACTGCTGCTGAGCAATTCACATGAAGAGGTTGAGAAGCT 1246
 QY 1306 TGTGCAACCTGCTGCTGAGCTGAGCTGCTGAGCAATTCACATGAAGAGGTTGAGAAGCT 1365
 DB 1247 TGTGCAACCTGCTGCTGAGCTGAGCTGCTGAGCAATTCACATGAAGAGGTTGAGAAGCT 1306
 QY 1366 TCTGTGAGCTTAACTACATGCTGTGTGGATATCACACTGAGTGTGCTGCTGACAG 1425
 DB 1307 TCTGTGAGCTTAACTACATGCTGTGTGGATATCACACTGAGTGTGCTGCTGACAG 1366
 QY 1426 TGTCAAGCAGAAAGACTTTGTGCTCATGCTGTGTCTGAAGAAACGATGGGGAACCT 1485
 DB 1367 TGTCAAGCAGAAAGACTTTGTGCTCATGCTGTGTCTGAAGAAACGATGGGGAACCT 1426
 QY 1486 TATGTGAGCAGACATCCACAGTGAAGAGGATATGCTCTTCTTCTTCTGATCT 1545
 DB 1427 TATGTGAGCAGACATCCACAGTGAAGAGGATATGCTCTTCTTCTTCTGATCT 1486
 QY 1546 TCTGTGCTGGGAGACTTCAAGACATTTGTGGCTGAGAGGCTATTAAAGACACACAT 1605
 DB 1487 TCTGTGCTGGGAGACTTCAAGACATTTGTGGCTGAGAGGCTATTAAAGACACACAT 1546
 QY 1606 ATCAGTGAATTAATCATTAACCTCCCTGTCACATCTTGCCCAATGGGGAATGATCT 1665
 DB 1547 ATCAGTGAATTAATCATTAACCTCCCTGTCACATCTTGCCCAATGGGGAATGATCT 1606
 QY 1666 TTACCAAAAGAGCTACACAGCATTTTCCACAGAGATGC 1703
 DB 1607 TTACCAAAAGAGCTACACAGCATTTTCCACAGAGATGC 1644

RESULT 2
 AA220792
 ID AA220792 standard; DNA; 2032 BP.
 XX
 XX AA220792;
 DF 08-DEC-1999 (first entry)
 XX Human glycosyl sulfotransferase-3 coding sequence.
 DE Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
 XX selectin binding interaction; inflammation; lymphocyte homing; human;
 KW

KW secondary lymph organ; ss.
 XX Homo sapiens.
 OS MO9949018-A1.
 XX 30-SEP-1999.
 XX 26-FEB-1999; 99WO-US04316.
 XX 20-MAR-1998; 98US-0045284.
 PR 12-NOV-1998; 98US-0190911.
 XX (REGC) UNIV CALIFORNIA.
 PA (SYNT) SYNTEX USA INC.
 XX Blstrup A, Rosen SD, Tangemann K, Hemmerich S;
 PI WPL; 1999-580442/49.
 DR P-PSDB; AAY3918.
 XX Human and murine glycosyl sulfotransferase 3 and related
 PT polynucleotides
 XX
 PS Claim 4; Fig 1; 59pp; English.
 CC This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.
 CC
 XX Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other:
 SO
 Query Match 75.7%; Score 1547; DB 20; Length 2032;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Oy 586 GCTCTGTGAGTCAAGACAGCCCTTTGAGTGTGTGAGAGAGGCGTGGCCCTCTCAAGCA 645
 |||||||
 Db 640 GCTCTGTGAGTCAAGACAGCCCTTTGAGTGTGTGAGAGAGGCGTGGCCCTCTCAAGCA 699
 |||||||
 Oy 646 CGTGGTGTCAAGAGAGGTGCGCTTCTTCACCTGAGTCCCTCAACCCGCTGTGAAAGA 705
 |||||||
 Db 700 CGTGGTGTCAAGAGAGGTGCGCTTCTTCACCTGAGTCCCTCAACCCGCTGTGAAAGA 759
 |||||||
 Oy 706 CCCCTCCCTCAACCTGATATGCTGACACTGTGTCGGGAGCCCGGGCCGCTTCGCTTC 765
 |||||||
 Db 760 CCCCTCCCTCAACCTGATATGCTGACACTGTGTCGGGAGCCCGGGCCGCTTCGCTTC 819
 |||||||
 Oy 766 CCGAGAACGCAAGAGAGAGATCTCATGATTGATGACAGTGCATTGTGATGGGCGAGCATGA 825
 |||||||
 Db 820 CCGAGAACGCAAGAGAGAGATCTCATGATTGATGACAGTGCATTGTGATGGGCGAGCATGA 879
 |||||||
 Oy 826 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTGTGAGGTCAATCTGCCAAAGCCA 885
 |||||||
 Db 880 GCAGAAACTCAAGAGAGAGACCAACCTTACTATGTGTGAGGTCAATCTGCCAAAGCCA 939
 |||||||
 Oy 886 GCTGGAGATCTACAAAGACATCCAGTCTCTTCCCAAGGCCCTGCAGAGAACCTACCTGCT 945
 |||||||
 Db 940 GCTGGAGATCTACAAAGACATCCAGTCTCTTCCCAAGGCCCTGCAGAGAACCTACCTGCT 999
 |||||||
 Oy 946 TGTGCGCTATGAGACCTGCGCTCGAGCCCTGTGGCCAGACTTCCGGAATGATGAAT 1005
 |||||||
 Db 1000 TGTGCGCTATGAGACCTGCGCTCGAGCCCTGTGGCCAGACTTCCGGAATGATGAAT 1059
 |||||||
 Oy 1006 CGTGGATTTGGAATTTCTTGCCCATCTTCAGACCTGGGTGATTAATCAATCAACCCGAGCA 1065
 |||||||
 Db 1060 CGTGGATTTGGAATTTCTTGCCCATCTTCAGACCTGGGTGATTAATCAATCAACCCGAGCA 1119
 |||||||
 Oy 1066 GGGCATGGGTGACCAAGCGTTTCCACACAAAGGCCAGGATCCCTTATGTCTCCAGGC 1125
 |||||||
 Db 1120 GGGCATGGGTGACCAAGCGTTTCCACACAAAGGCCAGGATCCCTTATGTCTCCAGGC 1179
 |||||||
 Oy 1126 TTGGCGCTGCTTTTGGCCCTATGAAAAGTTCTCGAATTCAGAAAAGCCTGTGGCGATGC 1185
 |||||||
 Db 1180 TTGGCGCTGCTTTTGGCCCTATGAAAAGTTCTCGAATTCAGAAAAGCCTGTGGCGATGC 1239
 |||||||
 Oy 1186 CATGAATTTGCTGGGCTTACCGCCACAGTGAATCTGAAACAAAGAGAAACCTGTGCT 1245
 |||||||
 Db 1240 CATGAATTTGCTGGGCTTACCGCCACAGTGAATCTGAAACAAAGAGAAACCTGTGCT 1299
 |||||||
 Oy 1246 GGATCTTCTGTCTACCTGAGCTGTCCTGAGCAAAATCCATGAAGGGTTGAGAAAGGCT 1305
 |||||||
 Db 1300 GGATCTTCTGTCTACCTGAGCTGTCCTGAGCAAAATCCATGAAGGGTTGAGAAAGGCT 1359
 |||||||
 Oy 1306 TGTGCGACCTGTGTGTCAGCTTCTCTGTAATGCTTCTGAGCCCTTGCTACA 1365
 |||||||
 Db 1360 TGTGCGACCTGTGTGTCAGCTTCTCTGTAATGCTTCTGAGCCCTTGCTACA 1419
 |||||||
 Oy 1366 TCTCTGAGCCTTAACTACATCTGTGTGGGTATCACACTGAGTGTGATGCTCCACAG 1425
 |||||||
 Db 1420 TCTCTGAGCCTTAACTACATCTGTGTGGGTATCACACTGAGTGTGATGCTCCACAG 1479
 |||||||
 Oy 1426 TGTCTCAAGCAGAAGACTTGTGTGTCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1485
 |||||||
 Db 1480 TGTCTCAAGCAGAAGACTTGTGTGTCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1539
 |||||||
 Oy 1486 TATGTGAGCAGACATCCACAGTAGTGAAGAGGGTATGTGCTTCTTCTTGTGATCT 1545
 |||||||
 Db 1540 TATGTGAGCAGACATCCACAGTAGTGAAGAGGGTATGTGCTTCTTCTTGTGATCT 1599
 |||||||
 Oy 1546 TCTGTGTGGGACAGCTTCAAGAGACTTGTGGCTGTGAGGCGCTTTTAAGCAGACAGT 1605
 |||||||
 Db 1600 TCTGTGTGGGACAGCTTCAAGAGACTTGTGGCTGTGAGGCGCTTTTAAGCAGACAGT 1659
 |||||||
 Oy 1606 ATCAGTGAATTTGATCCATTAATCCCTGTCACATCTTCCCAATGGGGAATGATCT 1665
 |||||||
 Db 1660 ATCAGTGAATTTGATCCATTAATCCCTGTCACATCTTCCCAATGGGGAATGATCT 1719
 |||||||
 Oy 1666 TTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703

DB 1720 TTACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1757

RESULT 3

AA516947 standard; cDNA; 1333 BP.

AA516947;

12-MAR-2002 (first entry)

Human L-selectin sulfotransferase-2 (LSST-2) cDNA.

Human; beta1,3gnt; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss; L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes; ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus; allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2; delayed-type hypersensitivity reaction; hyperplastic thymus; antiulcer; antiinflammatory; antiproliferative; antidiabetic; dermatological; antiallergic.

Homo sapiens.

Key CDS Location/Qualifiers
111..1253
/tag= a
/product= "Human LSST-2"

WO200185177-A1.

15-NOV-2001.

10-MAY-2001; 2001WO-US15452.

11-MAY-2000; 2000US-0569320.

(BURN-) BURNHAM INST.

Fukuda M, Yeh J, Hirakawa N;

WPI; 2002-075226/10.

P-PSDB; AA011274.

New enzyme, useful for modifying acceptor molecule, comprises an isolated L-selectin sulfotransferase-2 that directs expression of L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or intestinal G15NAC 6-sulfotransferase

Claim 19; Fig 4; 98pp: English.

The present invention provides a method of modifying an acceptor molecule by contacting the acceptor with an isolated beta1,3-N-acetylglucosaminyltransferase (beta1,3gnt) or an active fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The invention also provides a method of treating or preventing an L-selectin-mediated condition by reducing the expression or activity of a beta1,3gnt that directs expression of a MECA-79 antigen. This can be done by administering to the subject an oligosaccharide L-selectin antagonist that inhibits binding of L-selectin to a MECA-79 antigen, for example by administering antibody material that specifically binds beta1,3gnt, and/or a beta1,3gnt antisense nucleic acid molecule. L-selectin sulfotransferase-2 (LSST-2) also directs MECA-79 antigen expression. Alternatively, the expression or activity of LSST-2 or its active fragment can be reduced in combination with reducing the expression or activity of beta1,3gnt. The method is useful for treating L-selectin mediated conditions such as Crohn's disease and ulcerative colitis, inflammatory disorders of the skin such as allergic contact dermatitis, psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type hypersensitivity reactions, diabetes and hyperplastic thymus. This sequence represents cDNA encoding human LSST-2.

Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;

Query Match 61.8%; Score 1262; DB 24; Length 1333;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	107	AAAGTCTTCACCTTCAGCAGCAGATGCTACCTTAAAAAATGAAGCTCTGCTGTTTCG	166
DB	72	AAAGTCTTCACCTTCAGCAGCAGATGCTACCTTAAAAAATGAAGCTCTGCTGTTTCG	131
QY	167	GTTTCCAGATGGAGCATCTGGCTATCTTCCATGATGACAGCCACAAACATCAGCTCC	226
DB	132	GTTTCCAGATGGAGCATCTGGCTATCTTCCATGATGACAGCCACAAACATCAGCTCC	191
QY	227	CTGTCTATGAAGCAGCAGCCGAGCCGATGACAGCTGCTGTTCTCTTCGCGCTCT	286
DB	192	CTGTCTATGAAGCAGCAGCCGAGCCGATGACAGCTGCTGTTCTCTTCGCGCTCT	251
QY	287	GGCTCTCTTTTGTGGGGCAGCTTTTGGGACAGCCAGATGTTTCTACCTGATGAG	346
DB	252	GGCTCTCTTTTGTGGGGCAGCTTTTGGGACAGCCAGATGTTTCTACCTGATGAG	311
QY	347	CCGCGCTGGCAGCTGTGATGACCTTCAAGCAGCAGCCGCTGATGCTGCATGAGCT	406
DB	312	CCGCGCTGGCAGCTGTGATGACCTTCAAGCAGCAGCCGCTGATGCTGCATGAGCT	371
QY	407	GTCGGGATCTGATACGGGCGCTCTTCTTGTGACATGACGCTTTGATGCTTACATG	466
DB	372	GTCGGGATCTGATACGGGCGCTCTTCTTGTGACATGACGCTTTGATGCTTACATG	431
QY	467	GAACCTGGTCCCGGAGAGATGACAGCTTCAAGTGGGAGAACGCGGCGCTGTCT	526
DB	432	GAACCTGGTCCCGGAGAGATGACAGCTTCAAGTGGGAGAACGCGGCGCTGTCT	491
QY	527	TCGTGACCTGCTGTGATCATCATCCCAAGATGAATCATCCCGGCTCACTGACAG	586
DB	492	TCGTGACCTGCTGTGATCATCATCCCAAGATGAATCATCCCGGCTCACTGACAG	551
QY	587	CTCTGTGACATCAACAGCCCTTTGAGTGTGTGAGAAAGGCTGCCGCTCTACAGCC	646
DB	552	CTCTGTGACATCAACAGCCCTTTGAGTGTGTGAGAAAGGCTGCCGCTCTACAGCC	611
QY	647	GTCGTGCTCAAGAGAGTGGGCTTCTCAACCTGACATCCCTGACCGGTGGAAGAC	706
DB	612	GTCGTGCTCAAGAGAGTGGGCTTCTCAACCTGACATCCCTGACCGGTGGAAGAC	671
QY	707	CCCTCCCTCAACCTGATATGTCGACCTGTCCGGAGCCCGGCGCTGTCCGTTCC	766
DB	672	CCCTCCCTCAACCTGATATGTCGACCTGTCCGGAGCCCGGCGCTGTCCGTTCC	731
QY	767	CGAGAACGCAAAAGGAGATCTCATGATTTGACAGTCGATTTGTATGGGCGAGCTAG	826
DB	732	CGAGAACGCAAAAGGAGATCTCATGATTTGACAGTCGATTTGTATGGGCGAGCTAG	791
QY	827	CAAAAACCTAAGAGAGAGACCAACCTCTATGATGATGATGATGATGATGATGATG	886
DB	792	CAAAAACCTAAGAGAGAGACCAACCTCTATGATGATGATGATGATGATGATGATG	851
QY	887	CTGAGATGTACAGACATTCAGTCTTGTCCCAAGCCCTGACAGAGCTACCTGCTT	946
DB	852	CTGAGATGTACAGACATTCAGTCTTGTCCCAAGCCCTGACAGAGCTACCTGCTT	911
QY	947	GTCGCTATGAGAGCTGCTGTGAGCCCTGTGAGCCCTGTGAGCCCTGTGAGCCCT	1006
DB	912	GTCGCTATGAGAGCTGCTGTGAGCCCTGTGAGCCCTGTGAGCCCTGTGAGCCCT	971
QY	1007	GTCGCTATGAGAGCTTCTGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT	1066
DB	972	GTCGCTATGAGAGCTTCTGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCT	1031
QY	1067	GGCATGGTGTACACAGCTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGCT	1126
DB	1032	GGCATGGTGTACACAGCTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGCT	1091

QY 1127 TGGCGCTGCTTTGGCCCTATGAAAGTTTCTGACTTCAGAAAGCCTGTGGGATGCC 1186
 Db 1092 TGGCGCTGCTTTGGCCCTATGAAAGTTTCTGACTTCAGAAAGCCTGTGGGATGCC 1151
 QY 1187 ATGAATTTGCTGGGCTACCCGACAGTGTGACACAGAACAGAAACCTGTGCTG 1246
 Db 1152 ATGAATTTGCTGGGCTACCCGACAGTGTGACACAGAACAGAAACCTGTGCTG 1211
 QY 1247 GATCTTCTGCTACCTGAGCTGTGCTGACCAATTCACCTAAGAGGTTGAGAGGCTTT 1306
 Db 1212 GATCTTCTGCTACCTGAGCTGTGCTGACCAATTCACCTAAGAGGTTGAGAGGCTTT 1271
 QY 1307 GCTGCCACCTGCTGCTGAGCCTCAGTCACTTCTCTGATGCTTGTGAGCCTTGCTACAT 1366
 Db 1272 GCTGCCACCTGCTGCTGAGCCTCAGTCACTTCTCTGATGCTTGTGAGCCTTGCTACAT 1331
 QY 1367 CT 1368
 Db 1332 CT 1333

RESULT 4

AAZ94211
 ID AAZ94211 standard; cDNA; 2065 BP.

AC AAZ94211;

DT 19-JUN-2000 (first entry)

DE Human transferase TRNSFS-11 cDNA clone 2617407CBI.

XX Transferase; TRNSFS-11; human; antitumor; cell proliferation;

KM inflammation; gastrointestinal disorder; developmental disorder;

KM genetic disorder; neurological disorder; reproductive disorder;

KM smooth muscle disorder; immunological disorder; gene therapy;

XX diagnosis; N-acetylglicosamine 6-O-sulfotransferase; ss.

OS Homo sapiens.

FT Key location/Qualifiers

FT CDS 174..1334

PN MO200014251-A2.

PD 16-MAR-2000.

PE 09-SEP-1999; 99WO-US20989.

XX 10-SEP-1998; 98US-0150657.

PR 04-NOV-1998; 98US-0186779.

PR 11-MAY-1999; 99US-0133642.

PA (INCY-) INCYTE PHARM INC.

XX Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;

PI Hillman JL, Azimzal Y;

DR WPI: 2000-256996/22.

DR P-PSDB: AAY79219.

XX Human transferase proteins useful for preventing, diagnosing and

PT treating cancers and developmental, gastrointestinal, genetic,

PT immunological, neurological, reproductive and smooth muscle disorders -

Claim 9; Page 104-105; 113pp: English.

CC The present sequence is that of cDNA clone 2617407CBI encoding

CC human transferase TRNSFS-11 (see AAY79219). 1 of 15 claimed human

CC transferase proteins of the invention (see AAY79209-23). The clone

CC was isolated from gall bladder cDNA library GBLANOT01. TRNSFS-11

CC is expressed in dermatologic and gastrointestinal tissues,

CC especially those associated with inflammation and cell

CC proliferation. It shows homology to mouse N-acetylglicosamine
 CC 6-O-sulfotransferase. A polynucleotide comprising nucleotides
 CC 264-333 or 1272-1331 of the present sequence can be used as a DNA
 CC probe. The new human transferases and polynucleotides can be used
 CC in the diagnosis, prevention and treatment (including gene therapy
 CC and antisense therapy) of cancer, developmental disorders,
 CC gastrointestinal disorders, genetic disorders, immunological
 CC disorders, neurological disorders, reproductive disorders, and
 CC smooth muscle disorders.

Sequence 2065 BP; 466 A; 575 C; 491 G; 533 T; 0 other:

Query Match 56.4%; Score 1153; DB 21; Length 2065;

Best Local Similarity 99.6%; Pred. No. 0;

Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 106 CAGGCTTTCACCTTCAGCAGACATGCTACCTGCTTAAAAAATGAAGCTCCTGCTTTCT 165
 Db 152 CAGGCTTTCACCTTCAGCAGACATGCTACCTGCTTAAAAAATGAAGCTCCTGCTTTCT 211
 QY 166 GGTTCAGATGAGCATCTTGGCTTATCTTCACATGACAGCCAAACATCAGCTC 225
 Db 212 GGTTCAGATGAGCATCTTGGCTTATCTTCACATGACAGCCAAACATCAGCTC 271
 QY 226 CCTGCTATGAGGACAGACAGCCAGCCAGTACAGTGTGCTGCTTCTTCTGCGCTC 285
 Db 272 CCTGCTATGAGGACAGACAGCCAGCCAGTACAGTGTGCTGCTTCTTCTGCGCTC 331
 QY 286 TGGCTCTTCTTTTGGGGGAGCTTTTGGGAGACCCAGATGTTTCTACCTGATGA 345
 Db 332 TGGCTCTTCTTTTGGGGGAGCTTTTGGGAGACCCAGATGTTTCTACCTGATGA 391
 QY 346 GCGCCGCTGGAGCTGTGAGTACCTTCACAGCAGACAGCCCGTGAATGCTGACATGAGC 405
 Db 392 GCGCCGCTGGAGCTGTGAGTACCTTCACAGCAGACAGCCCGTGAATGCTGACATGAGC 451
 QY 406 TGTGCGGATCTGATAGGAGGCGCTTCTTGTGAGATGAGAGCTTTTGTATGCTACAT 465
 Db 452 TGTGCGGATCTGATAGGAGGCGCTTCTTGTGAGATGAGAGCTTTTGTATGCTACAT 511
 QY 466 GGAACCTGTGCTCCCGAGACAGTCCAGCTTTCAGTGGAGAAACAGCCGCGCTGTG 525
 Db 512 GGAACCTGTGCTCCCGAGAGACAGTCCAGCTTTCAGTGGAGAAACAGCCGCGCTGTG 571
 QY 526 TTTGCAACCTGCGTACATCATCCACAAGATGAATATCCCGGGGCTCATCTGAG 585
 Db 572 TTTGCAACCTGCGTACATCATCCACAAGATGAATATCCCGGGGCTCATCTGAG 630
 QY 586 GCTCTGTGAGTCAACAGCCCTTTG-AGTGTGTGAGAAAGGCTGCGCTCTACAGCC 644
 Db 631 GCTCTGTGAGTCAACAGCCCTTTGAGAGTTGTGAGAAAGGCTGCGCTCTACAGCC 690
 QY 645 ACGTGTGCTCAGAGAGGCTGCTTTCACCTGACAGTCCCTTACCCGCTGTAAG 704
 Db 691 ACGTGTGCTCAGAGAGGCTGCTTTCACCTGACAGTCCCTTACCCGCTGTAAG 750
 QY 705 ACCCGTCCCAACCTGATATGCTGACAGCTGCTGCGGAGACCCCGGGGCTGTTCCGTT 764
 Db 751 ACCCGTCCCAACCTGATATGCTGACAGCTGCTGCGGAGACCCCGGGGCTGTTCCGTT 810
 QY 765 CCCGAGAACGCAAAAGGAGATCTCATGATTGACAGTGCATGATGAGGAGCAGCATG 824
 Db 811 CCCGAGAACGCAAAAGGAGATCTCATGATTGACAGTGCATGATGAGGAGCAGCATG 870
 QY 825 AGCAAAATCTCAGAGAGGAGCAGACCTTACTATGTGATGACAGTATGCTGCCAAAGCC 884
 Db 871 AGCAAAATCTCAGAGAGGAGCAGACCTTACTATGTGATGACAGTATGCTGCCAAAGCC 930
 QY 885 AGCTGGAGATCTTAAGACATCCAGTCTGCGGAGAGCCGCTGAGAAAGCTACCTGTC 944
 Db 931 AGCTGGAGATCTTAAGACATCCAGTCTGCGGAGAGCCGCTGAGAAAGCTACCTGTC 990
 QY 945 TTGTGCGCTATGAGACCTGCTGAGACCCCTGTGCGCCAGACTTCCCGAATGTATGAT 1004

```

Db      991 TTGTGCGCTATGAGACCGCTGCGCCCTGTGGCCAGACTTCCCAATGATGAT 1050
Oy      1005 TCGTGGAGTTGGAAATCTTGGCCCATCTTCAGACCGTGGTGCATTAACATCACCCGAGGCA 1064
Db      1051 TCGTGGAGTTGGAAATCTTGGCCCATCTTCAGACCGTGGTGCATTAACATCACCCGAGGCA 1110
Oy      1065 AGGGCATGGGTGACCAACGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGG 1124
Db      1111 AGGGCATGGGTGACCAACGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAGG 1170
Oy      1125 CTTGGCGCTGTCTTGGCCCATCTTCAGACCGTGGTGCATTAACATCACCCGAGGCA 1184
Db      1171 CTTGGCGCTGTCTTGGCCCATCTTCAGACCGTGGTGCATTAACATCACCCGAGGCA 1230
Oy      1185 CCATGAATTTGGTGGCTACCGCCACAGTCAGATCTGACAAACAGAGAAACCTGTTC 1244
Db      1231 CCATGAATTTGGTGGCTACCGCCACAGTCAGATCTGACAAACAGAGAAACCTGTTC 1290
Oy      1245 TGGATCTTCTGTCTACCTGAGACTGTCCGAGCAAAATCCACTAAGAGGTTGAGAGGCT 1304
Db      1291 TGGATCTTCTGTCTACCTGAGACTGTCCGAGCAAAATCCACTAAGAGGTTGAGAGGCT 1350
Oy      1305 TTGCTGACCACTGTGTGACGCTCAGTCACTTCTGTGAATGCTTGTGAGCCCTGTGCTAC 1364
Db      1351 TTGCTGACCACTGTGTGACGCTCAGTCACTTCTGTGAATGCTTGTGAGCCCTGTGCTAC 1410
Oy      1365 ATCTGTGACCTTAACATCATCTGTGTGGTATCACACTGAGTGTGATGTGTCCACAC 1424
Db      1411 ATCTGTGACCTTAACATCATCTGTGTGGTATCACACTGAGTGTGATGTGTCCACAC 1470
Oy      1425 GTGCTCAACCAAGAGACTTTGTGTGCTGCTGTGTGTGTAAGAAACAGACGAGGAAAC 1484
Db      1471 GTGCTCAACCAAGAGACTTTGTGTGCTGCTGTGTGTGTAAGAAACAGACGAGGAAAC 1530
Oy      1485 TTATGTGAGCAGACATCCACCAAGTGAAGAGGATGTGTCTTCTTCTTCTGTATC 1544
Db      1531 TTATGTGAGCAGACATCCACCAAGTGAAGAGGATGTGTCTTCTTCTTCTGTATC 1590
Oy      1545 TTCTGTGTGTGGCAGACTTCAGACACTTGTGGCTGTGAGGCTTATTAAGACAGACAG 1604
Db      1591 TTCTGTGTGTGGCAGACTTCAGACACTTGTGGCTGTGAGGCTTATTAAGACAGACAG 1650
Oy      1605 TATCAGTGAATTCATTAACACCTGTGCTGCAATCTTGCCCAATGGGGAATGGATC 1664
Db      1651 TATCAGTGAATTCATTAACACCTGTGCTGCAATCTTGCCCAATGGGGAATGGATC 1710
Oy      1665 TTTCACCAAGAGCTCACAGCAATTTCCACAGAGATGC 1703
Db      1711 TTTCACCAAGAGCTCACAGCAATTTCCACAGAGATGC 1749

```

```

RESULT 5
AAK91803
ID      AAK91803 standard; cDNA; 877 BP.
AC      AAK91803;
XX      06-NOV-2001 (first entry)
DE      Human cDNA 5'-end sequence, SEQ ID NO: 263.
XX      Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX      Homo sapiens.
XX      EP130094-A2.
XX      05-SEP-2001.
XX      07-JUL-2000; 2000EP-0114089.
XX      08-JUL-1999; 99JP-0194486.

```

```

PR      11-JAN-2000; 2000JP-0118774.
PR      02-MAY-2000; 2000JP-0183765.
XX      (HELI-) HELIX RES INST.
XX      Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI      Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX      WPI; 2001-524255/58.
PT      830 Primers useful for synthesizing full length cDNA clones and their
PT      use in genetic manipulation.
XX      Claim 2; SEQ ID NO 263; 1380bp + sequence listing; English.
XX      The invention relates to primers for synthesizing full length cDNA
CC      clones. 830 cDNA molecules encoding a human protein have been
CC      isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC      molecules have been determined. Primers for synthesizing the full length
CC      cDNA are useful for clarifying the function of the protein encoded by
CC      the cDNA. The full length clones were obtained by construction of full
CC      length enriched cDNA libraries that were synthesized by the oligo-capping
CC      method. The primers enable the production of the full length cDNA easily
CC      without any special methods. The present sequence is the nucleotide
CC      sequence of the 5'-end of a cDNA provided in the invention.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in CD-ROM format directly from EPO.
XX      Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;
SQ
Query Match      34.0%; Score 695; DB 22; Length 877;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy      106 CAAGTCTTCCACTTTCAGACCAATGCTACTGCTTAAATAAATGAACTCTGCTTCTT 165
Db      47 CAAGTCTTCCACTTTCAGACCAATGCTACTGCTTAAATAAATGAACTCTGCTTCTT 106
Oy      166 GGTTCCTCCAGATGGCCATCTTGGCTATCTTCCACATGTACAGCCACAAATCAGCTC 225
Db      107 GGTTCCTCCAGATGGCCATCTTGGCTATCTTCCACATGTACAGCCACAAATCAGCTC 166
Oy      226 CCTGTATGAAGGACAGCCGAGGAGCAATGACAGTGTGTCTGCTTCCGAGGCTC 285
Db      167 CCTGTATGAAGGACAGCCGAGGAGCAATGACAGTGTGTGTCTGCTTCCGAGGCTC 226
Oy      286 TGGCTCTTCTTGTGTGGGAGAGCTTTTGGGAGACCCAGATGTTTCTACGTATGA 345
Db      227 TGGCTCTTCTTGTGTGGGAGAGCTTTTGGGAGACCCAGATGTTTCTACGTATGA 286
Oy      346 GCCCGCTGACAGCTGTGATGACCTTCAAGCAGACACCGCTGATGCTGACATGGC 405
Db      287 GCCCGCTGACAGCTGTGATGACCTTCAAGCAGACACCGCTGATGCTGACATGGC 346
Oy      406 TGTGGGAGATGATACGGGCGCTTCTTGTGCGCATGAGAGCTTTGTGATGCTACAT 465
Db      347 TGTGGGAGATGATACGGGCGCTTCTTGTGCGCATGAGAGCTTTGTGATGCTACAT 406
Oy      466 GGAACCTGTGCCCGGAGAGAGTCCAGCTCTTTCAGTGGAGAAACAGCCGGGCTGTG 525
Db      407 GGAACCTGTGCCCGGAGAGAGTCCAGCTCTTTCAGTGGAGAAACAGCCGGGCTGTG 466
Oy      526 TTCTGCACTGCTGTGACATCATCCACAGATGAATCATCCCGGGGCTCACTGCAG 585
Db      467 TTCTGCACTGCTGTGACATCATCCACAGATGAATCATCCCGGGGCTCACTGCAG 526
Oy      586 GCTCCTGTGACCAACAGCCCTTGAAGTGTGGAAGGCTGCGCTCTCTACAGCA 645
Db      527 GCTCCTGTGACCAACAGCCCTTGAAGTGTGGAAGGCTGCGCTCTCTACAGCA 586
Oy      646 CGTGTGTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTACCCGCTGTAAGA 705
Db      587 CGTGTGTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTACCCGCTGTAAGA 646

```

QY 706 CCCCCTCCCTAACCTGATATCGTGCACCTGTCGGGACCCCGGGCCGTTCCGTTTC 765
 DB 647 CCCCCTCCCTAACCTGATATCGTGCACCTGTCGGGACCCCGGGCCGTTCCGTTTC 706
 QY 766 CCGAGAACGCACAAAGGAGATCTCATGATTGACA 800
 DB 707 CCGAGAACGCACAAAGGAGATCTCATGATTGACA 741
 RESULT 6
 AAK93921
 ID AAK93921 standard; cDNA: 877 BP.
 AC AAK93921;
 XX
 XX
 DT 06-NOV-2001 (first entry)
 XX
 DE Human cDNA clone representative sequence, SEQ ID NO: 2381.
 XX
 KM Human, full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX
 XX Homo sapiens.
 PN EP130094-A2.
 XX
 XX 05-SEP-2001.
 PD 07-JUL-2000; 2000EP-0114089.
 XX
 PF 08-JUL-1999; 99JP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 XX (HELI-) HELIX RES INST.
 PA
 PI Oka T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 XX
 DR WPI; 2001-52425/58.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 XX Example 11; SEQ ID NO 2381; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence was used as the
 CC representative sequence from a human clone which was used in
 CC homology searches to identify the clone.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 SQ Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other:
 Query Match 34.0%; Score 695; DB 22; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 106 CAAGGCTTCACCTTCAGACCAATGCTAGTGCCTAAAAAATGAAGCTCTGCTGTTCT 165
 DB 47 CAAGGCTTCACCTTCAGACCAATGCTAGTGCCTAAAAAATGAAGCTCTGCTGTTCT 106
 QY 166 GTTTCCCAATGGCCATCTTGGCTATCTTCTTCACATGTACAGCCACAAATCAGCTC 225
 DB 107 GTTTCCCAATGGCCATCTTGGCTATCTTCTTCACATGTACAGCCACAAATCAGCTC 166

QY 226 CCGTCTATGAGGACACAGCCCGAGCGCATGCTGCTGTTCTTCTTCTGCGCTC 285
 DB 167 CCGTCTATGAGGACACAGCCCGAGCGCATGCTGCTGTTCTTCTTCTGCGCTC 226
 QY 286 TGGCTCTCTTTTGTGGGGCAGCTTTTGTGGGACAGCCACAGATGTTTCTACCTGATGA 345
 DB 227 TGGCTCTCTTTTGTGGGGCAGCTTTTGTGGGACAGCCACAGATGTTTCTACCTGATGA 286
 QY 346 GCCCGCTGCGACGTGTGATGACCTTTCACAGACAGCCGCTGATGCTGACATGGC 405
 DB 287 GCCCGCTGCGACGTGTGATGACCTTTCACAGACAGCCGCTGATGCTGACATGGC 346
 QY 406 TGGCGGGATCTGATPAGGGGCGCTTCTTGTGGACATGAGCGTCTTGTGCTACAT 465
 DB 347 TGTGGGGATCTGATPAGGGGCGCTTCTTGTGGACATGAGCGTCTTGTGATGCTACAT 406
 QY 466 GGAACCTGTGTCGCCGAGACAGTCCAGCTTCTTGTGAGGAGAACAGCCGCGCTGTG 525
 DB 407 GGAACCTGTGTCGCCGAGACAGTCCAGCTTCTTGTGAGGAGAACAGCCGCGCTGTG 466
 QY 526 TTCTGACCTGCTGTGACATCATCCACAAAGATGAATCATCCCGGGCTCACTGACG 585
 DB 467 TTCTGACCTGCTGTGACATCATCCACAAAGATGAATCATCCCGGGCTCACTGACG 526
 QY 586 GCTCTGTGACGTCAACAGCCCTTGTGAGGTGTGGAGAGGCGTCCGCTCTACAGCA 645
 DB 527 GCTCTGTGACGTCAACAGCCCTTGTGAGGTGTGGAGAGGCGTCCGCTCTACAGCA 586
 QY 646 CGTGTGCTCAAGAGGTGCGCTTCTTCAACCTGACGTCCCTTACCGCGTGTGAAGA 705
 DB 587 CGTGTGCTCAAGAGGTGCGCTTCTTCAACCTGACGTCCCTTACCGCGTGTGAAGA 646
 QY 706 CCCCCTCCCTCAACCTGATATGCTGACCTGTCCGGACCCCGGGCCGTTCCGTTTC 765
 DB 647 CCCCCTCCCTCAACCTGATATGCTGACCTGTCCGGACCCCGGGCCGTTCCGTTTC 706
 QY 766 CCGAGAACGCACAAAGGAGATCTCATGATTGACA 800
 DB 707 CCGAGAACGCACAAAGGAGATCTCATGATTGACA 741
 RESULT 7
 ABV89280
 ID ABV89280 standard; cDNA: 505 BP.
 XX
 AC ABV89280;
 XX
 DT 13-DEC-2002 (first entry)
 XX
 DE Human colon cancer related cDNA SEQ ID NO 2595.
 XX
 KM Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;
 XX ss.
 OS Homo sapiens.
 PN WO200258534-A2.
 XX
 PD 01-AUG-2002.
 PD
 PD 19-NOV-2001; 2001WO-US43704.
 PF
 PF 20-NOV-2000; 2000US-252222P.
 PR 06-FEB-2001; 2001US-267011P.
 PR 28-MAR-2001; 2001US-279670P.
 PR 10-JUL-2001; 2001US-304037P.
 PA (CORI-) CORIXA CORP.
 XX
 XX Stoik JA, Xu J, Chenaunt RA, Meagher MJ, Secrist H, King GE;
 PI WPI; 2002-608400/65.
 DR

XX New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer
 XX
 PS Claim 1: SEQ ID NO 2595; 266bp + Sequence listing; English.
 XX
 CC The invention relates to a human colon tumour expressed polynucleotide
 CC (I) encoding a polypeptide (II), ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV6669-ABV8289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (i) and (ii) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 505 BP; 120 A; 125 C; 119 G; 141 T; 0 other;

Query Match 19.7%; Score 403; DB 24; Length 505;

Best Local Similarity 100.0%; Pred. No. 2.6e-189; Mismatches 0; Gaps 0;

Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1301 GGCCTTGGTCCACACCTGCTGTCAGCCTTCTGTAATGCTTCTGAGCCTTGC 1360

DB 1 GGCCTTGGTCCACACCTGCTGTCAGCCTTCTGTAATGCTTCTGAGCCTTGC 60

DB 1361 CTACATCTGTGAGCCTTACTACATGCTGTGGGTATCAGCTGAGTGTGTCTC 1420

DB 61 CTACATCTGTGAGCCTTACTACATGCTGTGGGTATCAGCTGAGTGTGTCTC 120

DB 1421 ACACGTCGTCAGCAGCAAGAGCTTGTGTCATGCTTGTGCTGTAATAACAGACGGGG 1480

DB 121 ACACGTCGTCAGCAGCAAGAGCTTGTGTCATGCTTGTGCTGTAATAACAGACGGGG 180

DB 1481 AACCTTATGTGAGCAGCAGCATCCACAGTGAAGAGGATTTGCTTCTTCTTCTT 1540

DB 181 AACCTTATGTGAGCAGCAGCATCCACAGTGAAGAGGATTTGCTTCTTCTTCTT 240

DB 1541 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1600

DB 241 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 300

DB 1601 ACAGTATCAGTGAATGATCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGAATG 1660

DB 301 ACAGTATCAGTGAATGATCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGAATG 360

DB 1661 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1703

DB 361 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 403

DB 1661 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1703

DB 361 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 403

DB 1661 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1703

DB 361 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 403

DB 1661 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1703

DB 361 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 403

DB 1661 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1703

DB 361 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 403

DB 1661 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1703

DB 361 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 403

DB 1661 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1703

DB 361 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 403

DB 1661 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1703

XX 30-JUL-2001; 2001WO-US23826.

XX 03-AUG-2000; 2000US-223265P.

XX 02-OCT-2000; 2000US-237406P.

XX 20-MAR-2001; 2001US-277495P.

XX 03-JUL-2001; 2001US-302702P.

XX (COR-) CORIXA CORP.

XX Pyle RA, Xu J, Secretist H;

XX WPI, 2002-257462/30.

XX Novel polynucleotide encoding colon tumour polypeptides, useful as

XX vaccines for treating colon cancers

XX Claim 1; Page 206; 425bp; English.

XX The invention relates to isolated polynucleotides (I) encoding colon

XX tumour polypeptides (II). (I) is useful for stimulating an immune

XX response in a patient and treating colon cancer in a patient.

XX Oligonucleotides derived from (I) are useful for determining the presence

XX of cancer in a patient. (I) and (II) are useful in pharmaceutical

XX compositions, e.g. vaccines, and other compositions for the diagnosis

XX and treatment of colon cancer. A composition comprising a first component

XX selected from physiologically acceptable carriers and immunostimulants,

XX and an antigen-presenting cell expressing (II) is useful for inhibiting

XX development of cancer in a patient. (I) is useful in the design and

XX preparation of ribozyme molecules for inhibiting expression of tumour

XX polypeptides and (I). ABK54531-ABK54564 represent human colon cancer cDNA

XX sequences of the invention.

XX Sequence 517 BP; 121 A; 130 C; 124 G; 142 T; 0 other;

Query Match 19.7%; Score 403; DB 24; Length 517;

Best Local Similarity 100.0%; Pred. No. 2.6e-189; Mismatches 0; Gaps 0;

Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

DB 1301 GGCCTTGGTCCACACCTGCTGTCAGCCTTCTGTAATGCTTCTGAGCCTTGC 1360

DB 13 GGCCTTGGTCCACACCTGCTGTCAGCCTTCTGTAATGCTTCTGAGCCTTGC 72

DB 1361 CTACATCTGTGAGCCTTACTACATGCTGTGGGTATCAGCTGAGTGTGTCTC 1420

DB 73 CTACATCTGTGAGCCTTACTACATGCTGTGGGTATCAGCTGAGTGTGTCTC 132

DB 1421 ACACGTCGTCAGCAGCAAGAGCTTGTGTCATGCTTGTGCTGTAATAACAGACCTGGGG 1480

DB 133 ACACGTCGTCAGCAGCAAGAGCTTGTGTCATGCTTGTGCTGTAATAACAGACCTGGGG 192

DB 1481 AACCTTATGTGAGCAGCAGCATCCACAGTGAAGAGGATTTGCTTCTTCTTCTT 1540

DB 193 AACCTTATGTGAGCAGCAGCATCCACAGTGAAGAGGATTTGCTTCTTCTTCTT 252

DB 1541 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1600

DB 253 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 312

DB 1601 ACAGTATCAGTGAATGATCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGAATG 1660

DB 313 ACAGTATCAGTGAATGATCATTAACCTCCCTGTCACATCTTGGCCCAATGGGGAATG 372

DB 1661 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1703

DB 373 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 415

DB 1661 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1703

DB 373 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 415

DB 1661 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1703

DB 373 GATCTTCTGCTGTGGGACACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 415

XX 08-FEB-2001 (first entry)
 DT TTTGATGCTACATGGAACCTGTCCTCCGAGAGACAGTCCACGCTTTCAGTGGAGAAC 511
 XX 1480 TTTGATGCTACATGGAACCTGTCCTCCGAGAGACAGTCCACGCTTTCAGTGGAGAAC 1539
 DE Human OREF1711 polynucleotide sequence SEQ ID NO:3421.
 XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
 KW vulnary; antiparasitic; antiparkinsonian; neurotrophic; neuroprotective;
 KW anticonvulsant; osteopathic; antilarthritic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KW hypotensive; dermatological; immunosuppressive; antineoplastic;
 KW antiviral; antibacterial; antifungal; antihemetic; antihypertensive;
 KW antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive; ss.
 OS Homo sapiens.
 XX WO200058473-A2.
 XX 05-OCT-2000.
 PD 31-MAR-2000; 2000WO-US08621.
 PF 31-MAR-1999; 99US-0127607.
 XX 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX (CURA-) CURAGEN CORP.
 PA Shinkets RA, Leach M;
 PI WPI; 2000-602362/57.
 DR P-PSDB; ABA41947.
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX Claim 5; Page 2597-2599; 5507pp; English.
 PS AAC74446 to AAC77606 encode the proteins given in ABA40237 to ABA43397,
 CC which represent the human OREF open reading frames 1 to 3161. The OREF
 CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
 CC antiparasitic; antiparkinsonian; neurotrophic; neuroprotective;
 CC osteopathic; anticonvulsant; antilarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antibacterial; antiviral; antifungal; antihemetic;
 CC antihypertensive; and antineoplastic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an OREF-associated disorder. The
 CC nucleic acids can be used to express OREF proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other:
 SQ Query Match 15.0%; Score 306; DB 21; Length 2988;
 Best Local Similarity 100.0%; Pred. No. 3,7e-141;
 Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 TTTGATGCTACATGGAACCTGTCCTCCGAGAGACAGTCCACGCTTTCAGTGGAGAAC 511
 DB 1480 TTTGATGCTACATGGAACCTGTCCTCCGAGAGACAGTCCACGCTTTCAGTGGAGAAC 1539
 QY 512 AGCGGGCCCTGTGTTCTGACACCTGCTGACATCATCCACAAGATGAATCATCCCC 571
 DB 1540 AGCGGGCCCTGTGTTCTGACACCTGCTGACATCATCCACAAGATGAATCATCCCC 1599
 QY 572 CGGCTCTACTGAGGCTCTCTGACAGTCAACAGCCCTTGAAGGTGTGAGAGAGCCCTGC 631
 DB 1600 CGGCTCTACTGAGGCTCTCTGACAGTCAACAGCCCTTGAAGGTGTGAGAGAGCCCTGC 1659
 QY 632 CGGCTCTACTGAGGCTCTCTGACAGTCAACAGCCCTTGAAGGTGTGAGAGAGCCCTGC 691
 DB 1660 CGGCTCTACTGAGGCTCTCTGACAGTCAACAGCCCTTGAAGGTGTGAGAGAGCCCTGC 1719
 QY 692 CGGCTCTACTGAGGCTCTCTGACAGTCAACAGCCCTTGAAGGTGTGAGAGAGCCCTGC 751
 DB 1720 CGGCTCTACTGAGGCTCTCTGACAGTCAACAGCCCTTGAAGGTGTGAGAGAGCCCTGC 1779
 QY 752 GCCGTG 757
 DB 1780 GCCGTG 1785
 RESULT 10
 ID ABR54794
 ID ABR54794 standard; cDNA; 389 BP.
 AC ABR54794;
 XX 18-JUN-2002 (first entry)
 DE Human colon cancer-associated cDNA, SEQ ID NO 264.
 KW Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.
 OS Homo sapiens.
 XX WO200212280-A2.
 PN 14-FEB-2002.
 PD 30-JUL-2001; 2001WO-US23826.
 PF 03-AUG-2000; 2000US-223265P.
 PR 02-OCT-2000; 2000US-237406P.
 PR 20-MAR-2001; 2001US-277495P.
 PR 03-JUL-2001; 2001US-302702P.
 XX (CORI-) CORIXA CORP.
 PA Pyle RA, Xu J, Secretist R;
 PI WPI; 2002-257462/30.
 DR Novel polynucleotide encoding colon tumour polypeptides, useful as
 PT vaccines for treating colon cancers -
 XX Claim 1; Page 225; 425pp; English.
 PS The invention relates to isolated polynucleotides (I) encoding colon
 CC tumour polypeptides (II). (I) is useful for stimulating an immune
 CC response in a patient and treating colon cancer in a patient.
 CC Oligonucleotides derived from (I) are useful for determining the presence
 CC of cancer in a patient. (I) and (II) are useful in pharmaceutical
 CC compositions, e.g. vaccines, and other compositions comprising a first component
 CC and treatment of colon cancer. A composition comprising a first component
 CC selected from physiologically acceptable carriers and immunostimulants,
 CC and an antigen-presenting cell-expressing (II) is useful for inhibiting
 CC development of cancer in a patient. (I) is useful in the design and
 CC preparation of ribozyme molecules for inhibiting expression of tumour

CC polypeptides and (1). AAK54531-AAK5464 represent human colon cancer cDNA
 CC sequences of the invention.
 XX
 SQ Sequence 389 BP; 121 A; 83 C; 76 G; 109 T; 0 other;
 Query Match 5 6%; Score 114; DB 24; Length 389;
 Best Local Similarity 100.0%; Pred. No. 6.9e-46;
 Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1930 TGCACATCTGTAATAAATCCCAATTAAGTCTGTTAGATGTCCTTTTANG 1989
 DB 266 TGCACATCTGTAATAAATCCCAATTAAGTCTGTTAGATGTCCTTTTANG 325
 QY 1990 CTTCTTAATTATTAGCAGTAAGTTCATTTTANGGATCCCTAAAAA 2043
 DB 326 CTTCTTAATTATTAGCAGTAAGTTCATTTTANGGATCCCTAAAAA 379
 RESULT 11
 AAK92588/C
 ID AAK92588 standard; cDNA; 548 BP.
 AC AAK92588;
 XX
 DT 06-NOV-2001 (first entry)
 DE Human cDNA 3'-end sequence, SEQ ID NO: 1048.
 XX
 KM Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
 XX
 OS Homo sapiens.
 XX
 PN EP130094-A2.
 XX
 PD 05-SEP-2001.
 XX
 PF 07-JUL-2000; 2000EP-0114089.
 XX
 PR 08-JUL-1999; 99UP-0194486.
 PR 11-JAN-2000; 2000JP-0118774.
 PR 02-MAY-2000; 2000JP-0183765.
 XX
 PA (HELI-) HELIX RES INST.
 XX
 PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
 PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
 PI WPI; 2001-524255/58.
 XX
 PT 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 PS
 PS Claim 3: SEQ ID NO 1048; 1380bp + sequence listing; English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is the nucleotide
 CC sequence of the 3'-end of a cDNA provided in the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 CC
 XX
 SQ Sequence 548 BP; 160 A; 109 C; 129 G; 148 T; 2 other;
 Query Match 5.1%; Score 105; DB 22; Length 548;
 Best Local Similarity 100.0%; Pred. No. 2e-41;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGCACATCTGTAATAAATCCCAATTAAGTCTGTTAGATGTCCTTTTANG 1989
 DB 108 TGCACATCTGTAATAAATCCCAATTAAGTCTGTTAGATGTCCTTTTANG 49
 QY 1990 CTTCTTAATTATTAGCAGTAAGTTCATTTTANGGATCCCTAA 2034
 DB 48 CTTCTTAATTATTAGCAGTAAGTTCATTTTANGGATCCCTAA 4
 RESULT 12
 AAD24670
 ID AAD24670 standard; cDNA; 1647 BP.
 AC AAD24670;
 XX
 DT 12-MAR-2002 (first entry)
 DE Human drug metabolizing enzyme (DME)-5 cDNA.
 XX
 KM Human; drug metabolizing enzyme; gene therapy; autoimmune disorder;
 KM inflammatory disorder; acquired immune deficiency syndrome; infection;
 KM AIDS; adult respiratory distress syndrome; cell proliferative disorder;
 KM allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
 KM cancer; endocrine disorder; hypotaiamus disorder; pituitary disorder;
 KM gastrointestinal disorder; metabolic disorder; developmental disorder;
 KM liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
 KM goitre; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
 KM DME-5; ss.
 XX
 KM Homo sapiens.
 XX
 OS
 XX
 FH Key Location/Qualifiers
 FT CDS 1..1188
 FT /tag- a
 FT /product- "Human DME-5 protein"
 FT sig_peptide 1..96
 FT /tag- b
 FT mat_peptide 97..1185
 FT /tag- c
 FT /product- "Human mature DME-5 protein #1"
 FT sig_peptide 1..105
 FT /tag- d
 FT mat_peptide 106..1185
 FT /tag- e
 FT /product- "Human mature DME-5 protein #2"
 FT
 FT WO200179468-A2.
 PN
 PN 25-OCT-2001.
 PD
 XX
 PF 12-APR-2001; 2001WO-US11869.
 XX
 PR 13-APR-2000; 2000US-197590P.
 PR 19-APR-2000; 2000US-198403P.
 PR 28-APR-2000; 2000US-200185P.
 PR 05-MAY-2000; 2000US-202234P.
 PR 11-MAY-2000; 2000US-203509P.
 XX
 PA (INCY-) INCYTE GENOMICS INC.
 XX
 PI Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM;
 PI Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS;
 PI Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;
 PI Au-Young J;
 XX
 DR WPI: 2002-066363/09.
 DR P-PSDB; AAE15438.
 XX
 PT Novel isolated human drug metabolizing enzymes referred as DME 1-10,
 PT useful for diagnosing, treating, or preventing disorders associated
 PT with aberrant expression of DME such as allergy, anemia, asthma,
 PT infertility -

Fri Aug 15 08:09:31 2003

us-09-645-078-1_2.rng

Page 13

Db 834 AGGCTGCCGCTCCTACAGCCACGTGGTGTCTCAAGAGAGTGGCGCTTCTTCAACCTGCAG 892

Search completed: August 14, 2003, 21:37:42
Job time : 545 secs

THIS PAGE BLANK (USTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:16:48 : Search time 3887 Seconds
(without alignments)
12774.388 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 2043
Sequence: 1 gaattccattgttgggta.....tgggattcctaaataaaaaa 2043

Scoring table: OLIGO_NTC
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size: 0

Total number of hits satisfying chosen parameters: 45562182

Minimum DB seq length: 15
Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database:

EST:
1: em_estba:*
2: em_esthum:*
3: em_estln:*
4: em_estmu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_est1:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vrt:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gss1:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	685	33.5	954	12	B1823850
2	579	28.3	583	9	BSM082755
3	534	26.1	668	9	AL709927
4	479	23.4	571	12	BM129080

c	5	322	15.8	593	12	BM969292
c	6	290	14.2	680	10	BE878439
c	7	288	14.1	669	13	BU684395
c	8	218	10.7	553	12	BM128370
c	9	218	10.7	553	12	BM128831
c	10	107	5.2	362	9	AM002418
c	11	105	5.1	417	9	AM572510
c	12	101	4.9	358	10	BF056840
c	13	74	3.6	419	9	AM572390
c	14	59	2.9	525	10	BF197521
c	15	59	2.9	620	10	BF197521
c	16	59	2.9	620	10	BF197521
c	17	59	2.9	620	10	BF197521
c	18	55	2.7	721	28	AO741911
c	19	47	2.3	436	9	AM081348
c	20	27	1.3	507	12	BM246681
c	21	27	1.3	571	12	BM245312
c	22	27	1.3	695	12	BG964671
c	23	27	1.3	783	12	BG963298
c	24	27	1.3	852	12	BG966340
c	25	27	1.3	1923	11	AK009113
c	26	23	1.1	392	10	BE340900
c	27	23	1.1	397	9	AI345721
c	28	23	1.1	687	14	CB167190
c	29	23	1.1	692	14	BY711808
c	30	23	1.1	834	13	BO884877
c	31	22	1.1	148	9	AI345073
c	32	22	1.1	160	13	BO255877
c	33	22	1.1	167	9	AI253971
c	34	22	1.1	167	9	AI253971
c	35	22	1.1	171	9	AI246069
c	36	22	1.1	187	9	AI251238
c	37	22	1.1	187	9	AI792899
c	38	22	1.1	191	9	AI251979
c	39	22	1.1	196	9	AI792857
c	40	22	1.1	253	9	AA076760
c	41	22	1.1	290	9	AI345395
c	42	22	1.1	361	9	AI311365
c	43	22	1.1	493	10	BF869376
c	44	22	1.1	501	28	BM287165
c	45	22	1.1	516	9	AA689056

ALIGNMENTS

RESULT 1
B1823850
LOCUS 954 bp mRNA linear EST 04-OCT-2001
DEFINITION 603039012P1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179826 5',
mRNA sequence.
ACCESSION B1823850
VERSION B1823850.1 GI:15935400
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 954)
AUTHORS NIH-MGC http://mgc.ncl.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished
COMMENT Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNLN)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNLN at:
http://image.llnl.gov
Plate: L14M1448 row: h column: 03
High quality sequence stop: 856.

FEATURES

Location/Qualifiers
1..954

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5179826"

/lab_host="DH10B"

/clone_11b="NH_MGC_115"

/note="Organ: pooled brain, lung, testis; Vector:

PCMV-SPORT6; Site_1: NotI; Site_2: EcoRV (destroyed); RNA

source anonymous pool of 6 male brains; age range 23-27; 1

male lung, age 27; and 1 male testis, age 69. Library is

oligo-dT primed and directionally cloned (EcoRV site is

destroyed upon cloning). Average insert size 1.8 kb.

Insert size range 1-3 kb. Library is normalized and

enriched for full-length clones and was constructed by C.

Gruber (Invitrogen). Research Genetics tracking code

021. Note: this is a NIH-MGC Library."

BASE COUNT 199 a 297 c 250 g 208 t

Query Match 33.5%; Score 685; DB 12; Length 954;

Best Local Similarity 100.0%; Pred. No. 1.1e-271;

Matches 685; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

51 GCGCATGCGCGCGCTGACAGTACGCTCTCAAAAGCAGGAGGAGCCCAAGCCACAGC 110
|||||
1 GCGCATGCGCGCGCTGACAGTACGCTCTCAAAAGCAGGAGGAGCCCAAGCCACAGC 60
111 TCTTCCACTTCAGCAAAATGCTACTGCTTAAATAATGAAGCCCTGCTTCTGCTT 170
|||||
61 TCTTCCACTTCAGCAAAATGCTACTGCTTAAATAATGAAGCCCTGCTTCTGCTT 120
|||||
171 CCCAGATGGCCATCTTGCTATCTTCCACATGACAGCCACAACATCAGCTCCCTGT 230
|||||
121 CCCAGATGGCCATCTTGCTATCTTCCACATGACAGCCACAACATCAGCTCCCTGT 180
|||||
231 CATTAAGGACAGCCGAGCGCATGACAGTCTGCTGCTTCTGCTTCTGCTGCTGCT 290
|||||
181 CATTAAGGACAGCCGAGCGCATGACAGTCTGCTGCTTCTGCTTCTGCTGCTGCT 240
|||||
291 CTTCTTTTGTGGGAGCTTTTGGGACAGCCAGCATGTTTCTACATGATGAGCCCG 350
|||||
241 CTTCTTTTGTGGGAGCTTTTGGGACAGCCAGCATGTTTCTACATGATGAGCCCG 300
|||||
351 CCTGCGACGTGTGATGACCTTCAAGACAGCAGCCGCTGATGCTGACATGCTGCTC 410
|||||
301 CCTGCGACGTGTGATGACCTTCAAGACAGCAGCCGCTGATGCTGACATGCTGCTC 360
|||||
411 GGGATCTGATGAGGCGGCTCTTGTGAGCATGAGGCTTGTGATGCTACATGAGAAC 470
|||||
361 GGGATCTGATGAGGCGGCTCTTGTGAGCATGAGGCTTGTGATGCTACATGAGAAC 420
|||||
471 CTGATCCCGGAGACAGTCCAGCTCTTCACTGAGGAGAACGCGGCGCTGCTGCTG 530
|||||
421 CTGATCCCGGAGACAGTCCAGCTCTTCACTGAGGAGAACGCGGCGCTGCTGCTG 480
|||||
531 CACCTGCGCTGTGATGACCTTCAAGACAGCAGCCGCTGATGCTGACATGCTGCTC 590
|||||
481 CACCTGCGCTGTGATGACCTTCAAGACAGCAGCCGCTGATGCTGACATGCTGCTC 540
|||||
591 TGTGAGTCAACAGCCCTTGAAGTGTGAGAGAGCCGCGCTGCTGACGACAGCTG 650
|||||
541 TGTGAGTCAACAGCCCTTGAAGTGTGAGAGAGCCGCGCTGCTGACGACAGCTG 600
|||||
651 TGTGAGTCAACAGCCCTTGAAGTGTGAGAGAGCCGCGCTGCTGACGACAGCTG 710
|||||
601 TGTGAGTCAACAGCCCTTGAAGTGTGAGAGAGCCGCGCTGCTGACGACAGCTG 660
|||||
711 CCTCAACCTGCATATCTGACACT 735
|||||
661 CCTCAACCTGCATATCTGACACT 685

RESULT 2

HS082755 standard; RNA; EST; 583 BP.

AC BX490456;

SV BX490456.1

XX 09-MAY-2003 (Rel. 75, Created)

DT 09-MAY-2003 (Rel. 75, Last updated, Version 1)

XX Homo sapiens mRNA; EST DKFZp686G0680_r1 (from clone DKFZp686G0680)

DE EST; expressed sequence tag.

XX Homo sapiens (human)

OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;

OC Eutheria; Primates; Catarrhini; Hominiidae; Homo.

XX [1]

RP 1-583

RA Ansoorge W., Krieger S., Regiert T., Rittmuller C., Schwager B.,

RA Mewes H.W., Well B., Amlid C., Osanger A., Robo G., Han M., Wiemann S.;

RT Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.

RL MIPS, Ingolstaedter Landstr.1, D-85764 Neuherberg, GERMANY

XX This is the 5' sequence of the clone insert.

CC Clone from S. Wiemann, Molecular Genome Analysis, German Cancer

CC Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;

CC sequenced by EMBL (European Molecular Biology Laboratory),

CC Heidelberg/Germany) within the cDNA sequencing consortium of

CC the German Genome Project.

CC No 51 sequence available.

CC This clone (DKFZp686G0680) is available at the RZPD in Berlin.

CC Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6,

CC 14055 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

XX Key Location/Qualifiers

FH source

FT 1..583

FT /db_xref="taxon:9606"

FT /mol_type="mRNA"

FT /organism="Homo sapiens"

FT /clone="DKFZp686G0680"

FT /clone_11b="686 (synonym: h1ccc3). Vector pSPORT1_Sfi; host

FT DH10B; sites SfiIA + SfiIB"

FT /dev_stage="adult"

FT /tissue_type="cDNA-collection"

XX Sequence 583 BP; 120 A; 172 C; 154 G; 137 T; 0 other;

SO Query Match 28.3%; Score 579; DB 2; Length 583;

Best Local Similarity 100.0%; Pred. No. 6.8e-228;

Matches 579; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

41 GAGGAGAAAAAGCGCATGCGCGCTGACAGTACGCTCTCAAAAGCAGGAGGAGCCCA 100
|||||
1 GAGGAGAAAAAGCGCATGCGCGCTGACAGTACGCTCTCAAAAGCAGGAGGAGCCCA 60
|||||
101 AGCCACAAGGTCTTCCACTTCAAGCAAAATGCTACGCTTCAAAAGCAGGAGGAGCCCA 160
|||||
61 AGCCACAAGGTCTTCCACTTCAAGCAAAATGCTACGCTTCAAAAGCAGGAGGAGCCCA 120
|||||
161 TTTTGTGTTTCCAGATGCGCATCTTGTGCTATCTTCCACATGATGACGACCAACATC 220
|||||
121 TTTTGTGTTTCCAGATGCGCATCTTGTGCTATCTTCCACATGATGACGACCAACATC 180
|||||
221 AGCTCCCTGTATGAGGACAGCCGAGGAGCATGACGCTGCTGCTGCTGCTGCTG 280
|||||
181 AGCTCCCTGTATGAGGACAGCCGAGGAGCATGACGCTGCTGCTGCTGCTGCTG 240
|||||

QY 281 CGCTCTGCTCTTCTTTTGTGGGACATTTTGGGACAGCCAGATGTTTCTACCTG 340
 |||||
 Db 241 CGCTCTGCTCTTCTTTTGTGGGACATTTTGGGACAGCCAGATGTTTCTACCTG 300
 |||||
 QY 341 ATGAGACCCCGCTGGACAGTGTGATGACCTTCAAGCAGACACCGCTGGATGCTGAC 400
 |||||
 Db 301 ATGAGACCCCGCTGGACAGTGTGATGACCTTCAAGCAGACACCGCTGGATGCTGAC 360
 |||||
 QY 401 ATGAGCTGTGGGGATCGATACGGGCGCTCTCTTGTGGCAGATGACGCTTTGATGCC 460
 |||||
 Db 361 ATGAGCTGTGGGGATCGATACGGGCGCTCTCTTGTGGCAGATGACGCTTTGATGCC 420
 |||||
 QY 461 TACATGACCTGTGCTCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAAACAGCCGAGCC 520
 |||||
 Db 421 TACATGACCTGTGCTCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAAACAGCCGAGCC 480
 |||||
 QY 521 CTGTGTTCTGACCTGCTGCTGTGACATCATCCACAAATGATCCCGGGCTGAC 580
 |||||
 Db 481 CTGTGTTCTGACCTGCTGCTGTGACATCATCCACAAATGATCCCGGGCTGAC 540
 |||||
 QY 581 TGCAGGCTCCTGTGCTGACATCAACAGCCCTTTGAGTGGTG 619
 |||||
 Db 541 TGCAGGCTCCTGTGCTGACATCAACAGCCCTTTGAGTGGTG 579
 |||||

RESULT 3
 AL709927 668 bp mRNA linear EST 12-JUN-2003
 LOCUS DKFP68602364_r1 686 (synonym: h1cc3) Homo sapiens cDNA clone
 AL709927
 AL709927.1 GI:19693282
 EST.
 Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 668)
 Bahr A., Lauber J., Mewes H.W., Well B., Amd C., Osanger A., Fob
 G., Han M. and Wismann S.
 EST (Bahr A., Lauber J., Mewes H.W., Well B., et al.)
 Unpublished
 CONTACT: Bahr A
 MIPs

INQolstaedter Landstr.1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email s.wiemann@dkfz-heidelberg.de;
 consortium of the German Genome Project.
 No sl sequence available.
 This clone (DKFP68602364) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 location/Qualifiers

FEATURES

source
 1. 668
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFP68602364"
 /tissue_type="human skeletal muscle"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_lib="686 (synonym: h1cc3)"
 /note="Vector: pTriplex2; Site_1: SfiIA; Site_2: SfiIB;
 cDNA-collection"
 BASE COUNT 132 a 202 c 178 g 154 t 2 others
 ORIGIN

Query Match 26.1%; Score 534; DB 9; Length 668;
 Best Local Similarity 100.0%; Freq. No. 2.3e-209;
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 AGAGAGAAAAAGCGCATGGCCGCGCTAGCAGTGGACCTCTCAAAAGCAGGAGAGCC 99
 |||||
 Db 6 AGAGAGAAAAAGCGCATGGCCGCGCTAGCAGTGGACCTCTCAAAAGCAGGAGAGCC 65
 |||||
 QY 100 AAGCACAAGGCTCTTCCATTCACATTCAGCAGATGCTACTGCTTAAAAAATGAAGCTCTGCT 159
 |||||
 Db 66 AAGCACAAGGCTCTTCCATTCACATTCAGCAGATGCTACTGCTTAAAAAATGAAGCTCTGCT 125
 |||||
 QY 160 GTTCTCTGTTCCCAAGATGGCAGTCTGGCTCTATTCTTCACATGTACAGCCAAACAT 219
 |||||
 Db 126 GTTCTCTGTTCCCAAGATGGCAGTCTGGCTCTATTCTTCACATGTACAGCCAAACAT 185
 |||||
 QY 220 CAGCTCCCTGCTATGTAAGGACAGCCGAGCAGTCCAGTGGCTGCTGCTCTGCTGCTG 279
 |||||
 Db 186 CAGCTCCCTGCTATGTAAGGACAGCCGAGCAGTCCAGTGGCTGCTGCTGCTGCTGCTG 245
 |||||
 QY 280 GCGCTGTGCTCTTCTTTGTGGGGACGCTTTTGGGACAGCCAGATGTTTCTACT 339
 |||||
 Db 246 GCGCTGTGCTCTTCTTTGTGGGGACGCTTTTGGGACAGCCAGATGTTTCTACT 305
 |||||
 QY 340 GATGAGAGCCGCGCTGGGACAGTGGATGATGACCTTCAAGCAGAGCAGCCGCTGGATGCTGCA 399
 |||||
 Db 306 GATGAGAGCCGCGCTGGGACAGTGGATGATGACCTTCAAGCAGAGCAGCCGCTGGATGCTGCA 365
 |||||
 QY 400 CATGCTGTGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 459
 |||||
 Db 366 CATGCTGTGGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 425
 |||||
 QY 460 CTACATGGAACCTGTGCTCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAAACAGCCGAGCC 519
 |||||
 Db 426 CTACATGGAACCTGTGCTCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAAACAGCCGAGCC 485
 |||||

RESULT 4
 BM129080 571 bp mRNA linear EST 12-MAR-2002
 LOCUS 1117c04.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 cDNA clone IMAGE:5676655 5' similar to TR:09Y3R3 09Y3R3
 N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE. ;, mRNA sequence.
 BM129080
 BM129080.1 GI:17123632
 EST.
 Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 571)
 Melton D., Brown J., Kenty G., Permutt A., Lee C., Kaestner K.,
 Lemishka I., Scarce M., Brestelli J., Gradwohl G., Clifton S.,
 Hillier L., Marra M., Pape D., Wylie T., Martin J., Blisstein A.,
 Schmitt A., Theisinger B., Ritter E., Ronko I., Bennett J., Cardenas
 M., Gibbons M., McCann R., Cole R., Tsagarishevili R., Williams T.,
 Jackson X. and Bowers Y.
 Endocrine Pancreas Consortium
 Unpublished
 Other ESTs: 1117c04.x1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@bioph.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Juliana Brown
 (brown@fas.harvard.edu) This sequence now available from the IMAGE
 consortium, for clone orders contact: info@image.llnl.gov

FEATURES High quality sequence stop: 434.

Location/Qualifiers
1. 571

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5676655"

/sex="Both"

/tissue_type="Islets of Langerhans"

/dev_stage="Adult"

/lab_host="DH10B"

/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"

/note="Organ: Pancreas; Vector: pSPORT1; Site: 1; Not 1; Site: 2; Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation; average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an Ecot of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT 109 a 169 c 143 g 150 t
ORIGINQuery Match 23 4%; Score 479; DB 12; Length 571;
Best Local Similarity 100.0%; Pred. No. 1.2e-186;

Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

106 CAGGCTCTTCACATTCAGCACAATGCTACTGCTTAAATAAGTACCTCTGCTTTCT 165
 73 CAGGCTCTTCACATTCAGCACAATGCTACTGCTTAAATAAGTACCTCTGCTTTCT 132
 166 GGTTCCTCCAGATGGCATTTGGCTCTATTCTTCCACATGTACGCCACAAATCAGCTC 225
 133 GGTTCCTCCAGATGGCATTTGGCTCTATTCTTCCACATGTACGCCACAAATCAGCTC 192
 226 CCGTCTATGAAGGACACAGCCGAGGAGCATGCGTGTCTCTTCCGCGGCGCTC 285
 193 CCGTCTATGAAGGACACAGCCGAGGAGCATGCGTGTCTCTTCCGCGGCGCTC 252
 286 TGGCTCTCTTTTGTGGGACAGCTTTTGGGACAGCACCAGATGTTTCTACCTGATGA 345
 253 TGGCTCTCTTTTGTGGGACAGCTTTTGGGACAGCACCAGATGTTTCTACCTGATGA 312
 346 GCCCGCTGGACAGCTGTGATGACCTTCAAGCAGACACCCGCTGATGCTGACATGGC 405
 313 GCCCGCTGGACAGCTGTGATGACCTTCAAGCAGACACCCGCTGATGCTGACATGGC 372
 406 TGTGGGAGATGATACGGGCGCTCTTGTGCGACATGAGGCTTTGATGCTACAT 465
 373 TGTGGGAGATGATACGGGCGCTCTTGTGCGACATGAGGCTTTGATGCTACAT 432
 466 GGAACCTGTCCCGGAGACAGTCCAGCTTTTCAAGTGGGAGAACAGCCGGGCGCTGTG 525
 433 GGAACCTGTCCCGGAGACAGTCCAGCTTTTCAAGTGGGAGAACAGCCGGGCGCTGTG 492
 526 TTCTGTGACCTGCTGTGACATCATCCCAAGAATGAATCATCCCCGGGCTTACATGCA 584
 493 TTCTGTGACCTGCTGTGACATCATCCCAAGAATGAATCATCCCCGGGCTTACATGCA 551

RESULT 5
 LOCUS BM969292 593 bp mRNA linear EST 20-FEB-2003
 DEFINITION UI-CF-ENO-acp-1-21-0-UI.s1 UI-CF-ENO Homo sapiens cDNA clone
 ACCESSION BM969292
 VERSION BM969292.1 GI:19586879

KEYWORDS

EST.
Homo sapiens (human)
Homo sapiens

SOURCE

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 593)

AUTHORS

Bonaldo,M.F., Lennon,G. and Soares,M.B.

TITLE

Normalization and subtraction: two approaches to facilitate gene discovery

JOURNAL

Genome Res. 6 (9), 791-806 (1996)

MEDLINE

97044477

PUBMED

8889548

COMMENT

Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=yes.

FEATURES

Location/Qualifiers

1. 593

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="UI-CF-ENO-acp-1-21-0-UI"

/tissue_type="Primary Lung Cystic Fibrosis Epithelial Cells"

/dev_stage="Adult"

/lab_host="DH10B (Life Technologies) (T1 phage resistant)"

/clone_lib="UI-CF-ENO"

/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a modified Polylinker; Site: 1; Ecot II; Site: 2; Not I; UI-CF-ENO is a cDNA library containing the following tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells. The library was constructed according to Bonaldo, Lennon and Soares 1996 Genome Research 6:791-806. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an Ecot I adaptor, digested with Not I, and cloned directionally into pT73-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGCT.

TAG LIB-UI-CF-ENO

TAG_TISSUE=Human Lung Epithelial Cell Lines untreated LPS

6hr to LPS 24h

TAG_SEQ=CTGCTCAGCT

BASE COUNT 157 a 127 c 137 g 172 t

ORIGIN

Query Match 15 8%; Score 322; DB 12; Length 593;
Best Local Similarity 100.0%; Pred. No. 6.6e-122;

Matches 322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1382 ACATGCTGTGGTATCACATGAGTGTGTTGTCACAGTCTCAACAGAGA 1441
 1442 CTTTGTGTTCATGCTTGTGTAGAAAACAGACTGGGAACCTTATGTGACACACAT 1501
 533 CTTTGTGTTCATGCTTGTGTAGAAAACAGACTGGGAACCTTATGTGACACACAT 474
 1502 CCCACAGTGAACAGAGGATATGCTCTTTCTTTTCTTGAATCTTCTGCTGGGACAGC 1561

```

|||||
Db 473 CCCACAGTGAAGAGGGATATGGCTCTTCTTTCTTGATCTCTGTCGGGACAGC 414
Oy 1562 TTCAGAGCTTTGGGCGCTGAGGCTTTATAGCAGCAGACAGATGATGTAATGTC 1621
Db 413 TTCAGAGCTTTGGGCGCTGAGGCTTTATAGCAGCAGATGATGTAATGTAATGTC 354
Oy 1622 CATAAAGCTCCCTGTCACATCTTGGCCCAATGGGAAATGATCTTACCAAGAGCTCA 1681
Db 353 CATAAAGCTCCCTGTCACATCTTGGCCCAATGGGAAATGATCTTACCAAGAGCTCA 294
Oy 1682 CCAGCATTTCCACAGAGATGC 1703
Db 293 CCAGCATTTCCACAGAGATGC 272

RESULT 6
BF878439 680 bp mRNA linear EST 17-JAN-2001
LOCUS MRO-ET0109-191100-002-H06 ET0109 Homo sapiens cDNA, mRNA sequence.
ACCESSION BF878439
VERSION BF878439.1 GI:12268569
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 680)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.
Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663
10737800
COMMENT Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?l=MR0&t2=MR0-ET0109-
191100-002-h06&t3=2000-11-19&t4=1)
Seq primer: puc 18 forward
High quality sequence start: 17
High quality sequence stop: 641.
Location/Qualifiers
1. 680
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/dev_stage="Adult"
/clone_1lb="ET0109"
/note="Organ: lung,tumor; Vector: puc18; Site:1; SmaI:
Site:2; SmaI: A mini-library was made by cloning products
derived from ORESTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research).
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
BASE COUNT 168 a 164 c 163 g 184 t 1 others
ORIGIN
Query Match 14.2%; Score 290; DB 10; Length 680;
Best Local Similarity 100.0%; Pred. No. 9.9e-109;

```

```

Matches 290; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 1414 TGTGTCCACAGTGTGTCAGAGAGAGCTTTGTGTCATGCTGTCTAGAAACAG 1473
Db 106 TGTGTCCACAGTGTGTCAGAGAGAGCTTTGTGTCATGCTGTCTAGAAACAG 165
Oy 1474 ACTGGGAACCTTATAGTAGAGACACATCCACAGTAACAGAGGATATGCTCTTC 1533
Db 166 ACTGGGAACCTTATAGTAGAGACACATCCACAGTAACAGAGGATATGCTCTTC 225
Oy 1534 TTTCTGTATCTTCTCTGCTGGGACAGCTTCAGAGACTTTGTGCTGGAGCCATTAA 1593
Db 226 TTTCTGTATCTTCTCTGCTGGGACAGCTTCAGAGACTTTGTGCTGGAGCCATTAA 285
Oy 1594 GCACGACAGATACAGTACGATGATATGATACATTAACCTCCGTCCACATCTGCCCATG 1653
Db 286 GCACGACAGATACAGTACGATGATATGATACATTAACCTCCGTCCACATCTGCCCATG 345
Oy 1654 GGAATGATCTTTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
Db 346 GGAATGATCTTTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 395

RESULT 7
BU684395/c 669 bp mRNA linear EST 07-OCT-2002
LOCUS UI-CF-ENO-aco-f-08-0-UI.s1 UI-CF-ENO Homo sapiens cDNA clone
DEFINITION UI-CF-ENO-aco-f-08-0-UI 3', mRNA sequence.
ACCESSION BU684395
VERSION BU684395.1 GI:23537302
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 669)
Bonaldi,M.F., Lennon,G. and Soares,M.B.
Normalization and subtraction: two approaches to facilitate gene
discovery
Genome Res. 6 (9), 791-806 (1996)
97044477
8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue procurement: Dr. M. J. Welsh, University of Iowa
cDNA library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLY-A-Tes.
Location/Qualifiers
1. 669
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone_1lb="UI-CF-ENO"
/note="Organ: Lung; Vector: pRTT3-Pac (Pharmacia) with a
modified polylinker; Site:1; EcoR I; Site:2; Not I;
UI-CF-ENO is a cDNA library containing the following
tissue(s): Primary lung Cystic Fibrosis Epithelial Cells."
FEATURES
source

```


JOURNAL COMMENT Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Department of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biohpc.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: Juliana Brown
(brownj@fas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 425.

FEATURES
source

1. .553
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:567655"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_id="Melton Normalized Human Islet 4 N4-His 1"
/note="Organ: Pancreas; Vector: pSPORT1; Site.1: Not 1;
Site.2: Sal 1; Starting library constructed using
SuperScript Plasmid Library Kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected, by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 163 a 111 c 127 g 152 t
ORIGIN

Query Match 10.7%; Score 218; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.3e-79;
Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1486 TATGTGAGCAGCAGATCCACAGTGAACAGGATATGCTCTTCTTTCTTGATCT 1545
DB 553 TATGTGAGCAGCAGATCCACAGTGAACAGGATATGCTCTTCTTTCTTGATCT 494
QY 1546 TCCTGCTGGGAGACTTCAGAGACTTGTGGCCTGAGGAGCTTAAAGCAGCAGT 1605
DB 493 TCCTGCTGGGAGACTTCAGAGACTTGTGGCCTGAGGAGCTTAAAGCAGCAGT 434
QY 1606 ATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTGCCCAATGGGAATGATCT 1665
DB 433 ATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTGCCCAATGGGAATGATCT 374
QY 1666 TTCACCAAGAGCTCAGCAGCATTTTCCACAGAGATGC 1703
DB 373 TTCACCAAGAGCTCAGCAGCATTTTCCACAGAGATGC 336

RESULT 10
AM002418/c 362 bp mRNA linear EST 27-OCT-1999
LOCUS
DEFINITION
W61903.x1 NCI_CGAP_G66 Homo sapiens cDNA clone IMAGE:2524564 3',
mRNA sequence.
ACCESSION
AM002418
VERSION
AM002418.1 GI:5849334
KEYWORDS
EST.
SOURCE
Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 362)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished

JOURNAL COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
cDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.

cDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrrp/image/image.html
Seq primer: -400P from Glibco.

FEATURES
source

1. .362
location/Qualifiers
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2524564"
/tissue_type="Pooled germ cell tumors"
/lab_host="DH10B"
/clone_id="NCI_CGAP_G66"
/note="Vector: pT73D-Pac (Pharmacia) with a modified
polylinker; Site.1: Not 1; Site.2: Eco RI; Plasmid DNA
from the normalized library NCI_CGAP_G64 was prepared, and
as circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
1257096-1259831, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 108 a 69 c 77 g 108 t
ORIGIN

Query Match 5.2%; Score 107; DB 9; Length 362;
Best Local Similarity 100.0%; Pred. No. 3.7e-33;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGCATCTGTAATACATAAATCCCAATAAGTCTCTTAATACCTCCTTTATG 1989
DB 107 TGCATCTGTAATACATAAATCCCAATAAGTCTCTTAATACCTCCTTTATG 48
QY 1990 CTTCCTAATTAATAGCAGTAATGTCATTTTATGGATCCTAATA 2036
DB 47 CTTCCTAATTAATAGCAGTAATGTCATTTTATGGATCCTAATA 1

RESULT 11
AM572510/c 417 bp mRNA linear EST 13-MAR-2000
LOCUS
DEFINITION
X618911.x2 NCI_CGAP_ucl1 Homo sapiens cDNA clone IMAGE:2751044 3',
mRNA sequence.
ACCESSION
AM572510
VERSION
AM572510.1 GI:7237243
KEYWORDS
EST.
SOURCE
Homo sapiens (human)

ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE
1 (bases 1 to 417)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index
Unpublished

JOURNAL COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/lresources.shtml
 Seq primer: -40UP from gibco
 High quality sequence stop: 416.
 Location/Qualifiers

FEATURES

source

1. 417
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2751044"
 /tissue_type="well-differentiated endometrial adenocarcinoma, 7 pooled tumors"
 /lab_host="DH10B"
 /clone_1lb="NCI-CGAP-Utl1"
 /note="Organ: uterus; Vector: pCMV-SPORT6, Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.75 kb. Life Technologies catalog #: 11538-014"
 BASE COUNT 140 a 65 c 80 g 132 t
 ORIGIN

Query Match 5.1%; Score 105; DB 9; Length 417;
 Best Local Similarity 100.0%; Pred. No. 2.3e-32;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1930 TCCCATCTGTAACTAAATCCCAATAGGTCCTTTAGATGCCCTTTTATG 1989
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 266 TCCCATCTGTAACTAAATCCCAATAGGTCCTTTAGATGCCCTTTTATG 207
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 1990 CTCTTAATTATTAGCAGTAATAGTTCATTTTATGATGCCAA 2034
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 206 CTCTTAATTATTAGCAGTAATAGTTCATTTTATGATGCCAA 162
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 12 358 bp mRNA 1linear EST 16-OCT-2000
 BF056840/c 7K10D12.x1 NCI-CGAP-GC6 Homo sapiens CDNA clone IMAGE:3443783 3',
 LOCUS mRNA sequence.
 DEFINITION

ACCESSION BF056840.1 GI:10810736
 VERSION BF056840
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 358)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgaps-r@mail.nih.gov
 Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: inf@image.llnl.gov
 Seq primer: -40UP from gibco.
 Location/Qualifiers

FEATURES

source

1. 358
 /organism="Homo sapiens"
 /mol_type="mRNA"

/db_xref="taxon:9606"
 /clone="IMAGE:3443783"
 /tissue_type="pooled germ cell tumors"
 /lab_host="DH10B"
 /clone_1lb="NCI-CGAP GC6"
 /note="Vector: p773D-Pac (Pharmacia) with a modified polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA from the normalized library NCI-CGAP-GC4 was prepared, and ss circles were made in vitro. Following HAP purification, this DNA was used as tracer in a subtractive hybridization reaction. The driver was PCR-amplified cDNAs from a pool of 5,000 clones made from the same library (clonoids 1257096-1258631, 1469064-1470983, and 1475592-1476743). Subtraction by Bento Soares and M. Fatima Bonaldo."
 BASE COUNT 107 a 68 c 82 g 101 t
 ORIGIN

Query Match 4.9%; Score 101; DB 10; Length 358;
 Best Local Similarity 100.0%; Pred. No. 1.1e-30;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1932 CCATCTGTAACTAAATCCCAATAGGTCCTTTAGATGCCCTTTTATGCT 1991
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 101 CCATCTGTAACTAAATCCCAATAGGTCCTTTAGATGCCCTTTTATGCT 42
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 OY 1992 TCTTAATTATTAGCAGTAATAGTTCATTTTATGATGCCATCCT 2032
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 41 TCTTAATTATTAGCAGTAATAGTTCATTTTATGATGCCATCCT 1

RESULT 13 419 bp mRNA 1linear EST 13-MAR-2000
 AM572390/c XU09A06.x2 NCI-CGAP-C014 Homo sapiens CDNA clone IMAGE:2799634 3',
 LOCUS mRNA sequence.
 DEFINITION

ACCESSION AM572390.1 GI:7237123
 VERSION AM572390
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens

REFERENCE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 1 (bases 1 to 419)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/lresources.shtml
 Seq primer: -40UP from gibco
 High quality sequence stop: 343.
 Location/Qualifiers

JOURNAL Contact: Robert Strausberg, Ph.D.
 COMMENT Email: cgaps-r@mail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima Bonaldo, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL, send email to: inf@image.llnl.gov
 Seq primer: -40UP from gibco
 High quality sequence stop: 343.
 Location/Qualifiers

FEATURES

source

1. 419
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2799634"
 /tissue_type="moderately-differentiated adenocarcinoma"
 /lab_host="DH10B"
 /clone_1lb="NCI-CGAP-C014"
 /note="Organ: colon; Vector: pCMV-SPORT6, Site_1: SalI; Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt. Average insert size 1.7 kb. Life Technologies catalog #: 11531-019"
 BASE COUNT 122 a 78 c 89 g 130 t

ORIGIN

Query Match 3.6%; Score 74; DB 9; Length 419;

Best Local Similarity 100.0%; Pred. No. 1.4e-19;

Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1630 TCCCTGTCACATCTTCCCAATGGGATGATGATCTTTCACCAAGAGCTCACCAGATT.1689
|||||
Db 419 TCCTGTCCACATCTTCCCAATGGGATGATGATCTTTCACCAAGAGCTCACCAGATT 360

QY 1690 TTCACAGAGATGC 1703
|||||
Db 359 TTCACAGAGATGC 346

RESULT 14
LOCUS BF197521/c 525 bp mRNA linear EST 03-NOV-2000
DEFINITION 768408.x1 NCI-CGAP_Kid11 Homo sapiens CDNA clone IMAGE:3642903 3'
similar to TR:Q9UBX3 Q9UBX3 N-ACETYLGLUCOSAMINE
6-O-SULFOTRANSFERASE.; mRNA sequence.

ACCESSION BF197521 GI:11086670
VERSION BF197521
KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 525)
Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL, send email to:

info@image.llnl.gov

High quality sequence stop: 451.

Location/Qualifiers

1..525

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:3642903"

/lab_host="DH10B"

/clone_1lb="NCI-CGAP_Kid11"

/note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with

a modified polylinker; Site1: Not I; Site2: Eco RI;

Plasmid DNA from the normalized library NCI-CGAP_Kid3

prepared, and ss circles were made in vitro. Following HAP

purification, this DNA was used as tracer in a subtractive

hybridization reaction. The diver was PCR-amplified cDNAs

from a pool of 5,000 clones made from the same library

(cloneids 1322376-1323911, 1456007-1456775, and

1500352-1502853). Subtraction by Bento Soares and M.

Fatima Bonaldo.

BASE COUNT 90 a 163 c 187 g 85 t

ORIGIN

Query Match 2.9%; Score 59; DB 10; Length 525;

Best Local Similarity 100.0%; Pred. No. 2e-13;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCTCTCTACAGCAGTGTCGTCAGAGAGGAGGCTTCTCAACTGCAG 682
|||||
Db 260 AGGCTGCGCTCTCTACAGCAGTGTCGTCAGAGAGGAGGCTTCTCAACTGCAG 202

RESULT 15

AI824100/c

LOCUS AI824100 620 bp mRNA linear EST 21-DEC-1999

DEFINITION wj46c01.x1 NCI-CGAP_Lu19 Homo sapiens CDNA clone IMAGE:2405856 3'

similar to TR:075667 075667 DJ7116.4; contains PFR5, b2 PFR5
repetitive element.; mRNA sequence.

ACCESSION AI824100.1 GI:5444771

VERSION AI824100

KEYWORDS EST.

SOURCE Homo sapiens (human)

ORGANISM Homo sapiens

REFERENCE 1 (bases 1 to 620)

Mammalia; Euthera; Primates; Catarrhini; Homnidae; Homo.

NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.

National Cancer Institute, Cancer Genome Anatomy Project (CGAP),

Tumor Gene Index

JOURNAL Unpublished

COMMENT Contact: Robert Strausberg, Ph.D.

Email: cgaps-remail.nih.gov

Tissue Procurement: Christopher Moskalko, M.D., Ph.D., Michael R.

Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: M. Bento Soares, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.

DNA Sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LNL at:

www.bio.llnl.gov/bdrrp/image/image.html

Insert Length: 917 Std Error: 0.00

Seq primer: -400P from Gldco

High quality sequence stop: 490.

Location/Qualifiers

1..620

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2405856"

/tissue_type="squamous cell carcinoma, poorly

differentiated (4 pooled tumors, including primary and

metastatic)"

/dev_stage="adult"

/lab_host="DH10B (phage-resistant)"

/clone_1lb="NCI-CGAP_Lu19"

/note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a

modified polylinker; 1st strand cDNA was prepared from

pooled lung tumor tissue, and was then primed with a Not I

- oligo(dT) primer. Double-stranded cDNA was ligated to

Eco RI adaptors (Pharmacia), digested with Not I and

cloned into the Not I and Eco RI sites of the modified

pT73 vector. Library went through one round of

normalization. Library constructed by Bento Soares and M.

Fatima Bonaldo.

BASE COUNT 106 a 195 c 227 g 89 t 3 others

ORIGIN

Query Match 2.9%; Score 59; DB 9; Length 620;

Best Local Similarity 100.0%; Pred. No. 1.9e-13;

Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCTCTCTACAGCAGTGTCGTCAGAGAGGAGGCTTCTCAACTGCAG 682
|||||
Db 314 AGGCTGCGCTCTCTACAGCAGTGTCGTCAGAGAGGAGGCTTCTCAACTGCAG 256

Search completed: August 15, 2003, 00:44:17
Job time: 3888 secs

THIS PAGE BLANK (USPTO)

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 21:19:32 : Search time 142 Seconds

(Without alignments)
6350.322 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattccattgtgttggtga.....tgggattccttaaaaaaaaaa 2043

Scoring table: OLIGO_NUC

Searched: 569978 seqs, 220691566 residues

Word size : 0

Total number of hits satisfying chosen parameters: 1092396

Minimum DB seq length: 15

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :
Issued_Patents_MN:*
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/PCTUS.COMB.seq:*
6: /cgn2_6/ptodata/1/lna/backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1547	75.7	2032	3	US-09-045-284A-1
2	1547	75.7	2032	4	US-09-190-911-2
3	1153	56.4	2065	4	US-09-786-240-26
4	37	1.8	37	4	US-09-045-284A-7
5	37	1.8	37	4	US-09-190-911-7
6	19	0.9	3792	2	US-08-992-334-1
7	19	0.9	3792	2	US-08-302-752-1
8	19	0.9	5234	2	US-08-992-334-2
9	19	0.9	5234	2	US-08-302-752-2
10	19	0.9	6722	2	US-08-992-334-3
11	19	0.9	6722	2	US-08-302-752-3
12	18	0.9	23	2	US-08-778-494B-112
13	18	0.9	47	3	US-09-045-284A-8
14	18	0.9	47	3	US-09-190-911-8
15	18	0.9	150	4	US-09-439-313-320
16	18	0.9	150	4	US-09-352-616A-320
17	18	0.9	150	4	US-09-232-149A-320
18	18	0.9	201	3	US-09-030-956-78
19	18	0.9	201	3	US-09-030-607-78
20	18	0.9	201	4	US-09-439-313-78
21	18	0.9	201	4	US-09-352-616A-78
22	18	0.9	201	4	US-09-232-149A-78
23	18	0.9	202	3	US-09-030-607-180
24	18	0.9	202	4	US-09-439-313-180
25	18	0.9	202	4	US-09-352-616A-180
26	18	0.9	202	4	US-09-232-149A-180
27	18	0.9	536	3	US-09-020-956-69

28	18	0.9	536	3	US-09-030-607-69	Sequence 69, App1
29	18	0.9	536	4	US-09-439-313-69	Sequence 69, App1
30	18	0.9	536	4	US-09-352-616A-69	Sequence 69, App1
31	18	0.9	536	4	US-09-232-149A-69	Sequence 69, App1
32	18	0.9	699	4	US-09-107-532A-1333	Sequence 29, App1
33	18	0.9	822	3	US-09-020-956-29	Sequence 29, App1
34	18	0.9	822	3	US-09-030-607-29	Sequence 29, App1
35	18	0.9	822	4	US-09-439-313-29	Sequence 29, App1
36	18	0.9	822	4	US-09-352-616A-29	Sequence 29, App1
37	18	0.9	822	4	US-09-232-149A-29	Sequence 29, App1
38	18	0.9	1197	4	US-09-257-179-26	Sequence 61, App1
39	18	0.9	1368	4	US-09-118-554-61	Sequence 61, App1
40	18	0.9	1368	4	US-09-118-627-61	Sequence 61, App1
41	18	0.9	1368	4	US-09-602-877A-61	Sequence 40, App1
42	18	0.9	1913	4	US-09-599-360B-40	Sequence 40, App1
43	18	0.9	1927	4	US-09-336-536-66	Sequence 66, App1
44	18	0.9	4351	4	US-09-634-238-25	Sequence 25, App1
45	17	0.8	141	1	US-08-036-555B-161	Sequence 161, App1

ALIGNMENTS

RESULT 1									
US-09-045-284A-1									
: Sequence 1, Application US/09045284A									
: Patent No. 6265192									
: GENERAL INFORMATION:									
: APPLICANT: Bistrup, Annette									
: APPLICANT: Rosen, Steven D.									
: APPLICANT: Hemmerich, Stefan									
: TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3									
: FILE REFERENCE: 6510-107U51									
: CURRENT APPLICATION NUMBER: US/09/045,284A									
: CURRENT FILING DATE: 1998-03-20									
: NUMBER OF SEQ ID NOS: 9									
: SOFTWARE: FastSeq for Windows Version 4.0									
: SEQ ID NO 1.									
: LENGTH: 2032									
: TYPE: DNA									
: ORGANISM: Homo sapiens									
US-09-045-284A-1									
Query Match									
Best Local Similarity 99.9%; Pred. No. 0;									
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;									
QY	106	CAAGGTCCTTCACCTTCAGACACATGCTACTGCTTAAAAAATGAGAGCTCTGCTTTCT	165						
DB	160	CAAGGTCCTTCACCTTCAGACACATGCTACTGCTTAAAAAATGAGAGCTCTGCTTTCT	219						
QY	166	GCTTTCACCATGGCCATCTTGGCTCTATTCTTCCACATGTACAGCCAAACATCAGCTC	225						
DB	220	GCTTTCACCATGGCCATCTTGGCTCTATTCTTCCACATGTACAGCCAAACATCAGCTC	279						
QY	226	CCTGTCTATGAGGACACAGCCGAGCGATGCTGCTGTCTTCTTCTGCGGCTC	285						
DB	280	CCTGTCTATGAGGACACAGCCGAGCGATGCTGCTGTCTTCTTCTGCGGCTC	339						
QY	286	TGCTCTTCTTTTGTGGGGCAGCTTTTGGGAGCAGCCAGATGTTTTCTTACCTGATGA	345						
DB	340	TGCTCTTCTTTTGTGGGGCAGCTTTTGGGAGCAGCCAGATGTTTTCTTACCTGATGA	399						
QY	346	GCCGCGCTGCGACGCTGTGATGACCTTCAAGACAGACACCGCTGTGATGCTGCACATGGC	405						
DB	400	GCCGCGCTGCGACGCTGTGATGACCTTCAAGACAGACACCGCTGTGATGCTGCACATGGC	459						
QY	406	TGTGCGGAGATCTGATACGGGCGCTTCTTGTGACATGAGCGCTTTGATGCTTACAT	465						
DB	460	TGTGCGGAGATCTGATACGGGCGCTTCTTGTGACATGAGCGCTTTGATGCTTACAT	519						
QY	466	GGAACCTGTGCTCCCGGAGACAGTCCAGCTCTTTTCAGTGGGAGAAAGCCGGGCGCTGTG	525						

```
Db 520 GGAACCTGGTCCCGGAGACAGTCAGCTCTTTCACTGGGAGAACAGCCGGGCTCTGTG 579
QY 526 TTCTGCACTGCTGCTGACATCATCCACAAAGATGAATCATCCCGGGCTCATCGAG 585
Db 580 TTCTGCACTGCTGCTGACATCATCCACAAAGATGAATCATCCCGGGCTCATCGAG 639
QY 586 GCTCCCTGACAGTCAACAGCCCTTTAGAGTGGTGGAGAAAGGCTGCCCTCTACAGCCA 645
Db 640 GCTCCCTGACAGTCAACAGCCCTTTAGAGTGGTGGAGAAAGGCTGCCCTCTACAGCCA 699
QY 646 CGTGAGTCAAGAGAGAGTGGCTTTCTCAACCTGACAGTCCCTTACCCGCTGCTGAAGA 705
Db 700 CGTGAGTCAAGAGAGAGTGGCTTTCTCAACCTGACAGTCCCTTACCCGCTGCTGAAGA 759
QY 706 CCCCTCCCTCAACCTGACATATCGTGCACCTGGTCCGGAGCCCGGGCGGTGTTCCGTTT 765
Db 760 CCCCTCCCTCAACCTGACATATCGTGCACCTGGTCCGGAGCCCGGGCGGTGTTCCGTTT 819
QY 766 CCGAGAACCCACAAAGAGAGATCTCATGATGACATCGCATTTGTGATGGGCGACATGA 825
Db 820 CCGAGAACCCACAAAGAGAGATCTCATGATGACATCGCATTTGTGATGGGCGACATGA 879
QY 826 GCAAAAACCTCAAGAGAGAGAGACCAACCTACTATGTAGACAGTGCATCGCCAAAGCCA 885
Db 880 GCAAAAACCTCAAGAGAGAGAGACCAACCTACTATGTAGACAGTGCATCGCCAAAGCCA 939
QY 886 GCTGAGATCTACAGAACCATCCAGTCTTGCCCAAGGCCCTGACAGAAAGCTTACCTGCT 945
Db 940 GCTGAGATCTACAGAACCATCCAGTCTTGCCCAAGGCCCTGACAGAAAGCTTACCTGCT 999
QY 946 TTGCGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1005
Db 1000 TTGCGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1059
QY 1006 CCGTGGATGGAATTTTCCCATCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1065
Db 1060 CCGTGGATGGAATTTTCCCATCTTACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1119
QY 1066 GGGATGAGTGAACACAGCTTTCACACAAATGCGAGAGAGAGAGAGAGAGAGAGAG 1125
Db 1120 GGGATGAGTGAACACAGCTTTCACACAAATGCGAGAGAGAGAGAGAGAGAGAGAG 1179
QY 1126 TTGGGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1185
Db 1180 TTGGGCTGAGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT 1239
QY 1186 CATGAATTTGCTGGGCTACCGCAGCTGAGATCTGAAGAAGAGAGAGAGAGAGAGAG 1245
Db 1240 CATGAATTTGCTGGGCTACCGCAGCTGAGATCTGAAGAAGAGAGAGAGAGAGAGAG 1299
QY 1246 GGAATCTTGTCTACCTGAGCTGCTGAGCAAAATCACTAAGAGAGAGAGAGAGAG 1305
Db 1300 GGAATCTTGTCTACCTGAGCTGCTGAGCAAAATCACTAAGAGAGAGAGAGAGAG 1359
QY 1306 TGTGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1365
Db 1360 TGTGCTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1419
QY 1366 TGTGAGAGCTTAACTACATGCTGTGGGTATCACAGTGAAGTGAAGTGAAGTGAAG 1425
Db 1420 TGTGAGAGCTTAACTACATGCTGTGGGTATCACAGTGAAGTGAAGTGAAGTGAAG 1479
QY 1426 TGTCAAGCAGAAAGACTTTTGTCTCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1485
Db 1480 TGTCAAGCAGAAAGACTTTTGTCTCAGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 1539
QY 1486 TATGAGAGCAGACATCCACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1545
Db 1540 TATGAGAGCAGACATCCACAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1599
QY 1546 TCCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1605
Db 1600 TCCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1659
```

```
QY 1606 ATCAGTGAATTTGATCCATTAACCTCCGTGCACATCTTGCCCAATGGGGAATGATCT 1665
Db 1660 ATCAGTGAATTTGATCCATTAACCTCCGTGCACATCTTGCCCAATGGGGAATGATCT 1719
QY 1666 TTACCAAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
Db 1720 TTACCAAAAGAGCTCACAGCATTTTCCACAGAGATGC 1757

RESULT 2
US-09-190-911-2
; Sequence 2, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerlich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; CURRENT APPLICATION NUMBER: US/09/190,911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-2

Query Match 75.7%; Score 1547; DB 4; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAAGTCTTCCACTTGAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTTCT 165
Db 160 CAAGTCTTCCACTTGAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTTCT 219
QY 166 GCTTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 225
Db 220 GCTTCCCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 279
QY 226 CCGTGTATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 285
Db 280 CCGTGTATGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 339
QY 286 TGGCTCTCTTTTGTGGGAGAGCTTTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 345
Db 340 TGGCTCTCTTTTGTGGGAGAGCTTTTGGGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 399
QY 346 GCGCGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 405
Db 400 GCGCGCTGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 459
QY 406 TGTGCGGATCTGATACGGGCGCTCTTGTGTGCGAGATGAGAGAGAGAGAGAGAGAGAG 465
Db 460 TGTGCGGATCTGATACGGGCGCTCTTGTGTGCGAGATGAGAGAGAGAGAGAGAGAGAG 519
QY 466 GGAACCTGGTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 525
Db 520 GGAACCTGGTCCCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 579
QY 526 TTCTGACCTGCTGCTGACATCATCCACAAAGATGAATCATCCCGGGCTCATCGAG 585
Db 580 TTCTGACCTGCTGCTGACATCATCCACAAAGATGAATCATCCCGGGCTCATCGAG 639
QY 586 GCTCCCTGACAGTCAACAGCCCTTTAGAGTGGTGGAGAAAGGCTGCCCTCTACAGCCA 645
Db 640 GCTCCCTGACAGTCAACAGCCCTTTAGAGTGGTGGAGAAAGGCTGCCCTCTACAGCCA 699
```


STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ANTI-SENSE: YES
IMMEDIATE SOURCE:
CLONE: pg-host4
US-08-992-334-1

Query Match 0.9%; Score 19; DB 2; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCTGTGACAAAGAGCAAG 1839
DB 1818 TCTGTGACAAAGAGCAAG 1836

RESULT 7
US-08-302-752-1
Sequence 1, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-1

Query Match 0.9%; Score 19; DB 3; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCTGTGACAAAGAGCAAG 1839
DB 1818 TCTGTGACAAAGAGCAAG 1836

RESULT 8
US-08-992-334-2
Sequence 2, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-2

Query Match 0.9%; Score 19; DB 2; Length 5234;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCCTGCAAAAGAGCAAG 1839
Db 3260 TCCTGCAAAAGAGCAAG 3278

RESULT 9

US-08-302-752-2
Sequence 2, Application US/08302752

Patent No. 6025190

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: THERMOSENSIBLE PLASMID

NUMBER OF SEQUENCES: 3

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (PPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/302,752

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 9203034

FILING DATE: 13-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: NO FR/93/00248

FILING DATE: 12-MAR-1993

INFORMATION FOR SEQ ID NO: 2:

SEQUENCE CHARACTERISTICS:

LENGTH: 5234 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-302-752-2

Query Match
Best Local Similarity 100.0%; Score 19; DB 3; Length 5234;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCCTGCAAAAGAGCAAG 1839
Db 3260 TCCTGCAAAAGAGCAAG 3278

RESULT 10

US-08-992-334-3
Sequence 3, Application US/08992334

Patent No. 5919678

GENERAL INFORMATION:

APPLICANT: Gruss, Alexandra

APPLICANT: Masulin, Emmanuelle

TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE

NUMBER OF SEQUENCES: 3

CORRESPONDENCE ADDRESS:

ADDRESSEE: Christie Parker & Hale, LLP

STREET: 350 West Colorado Boulevard, Suite 500

CITY: Pasadena

STATE: California

COUNTRY: United States

ZIP: 91105

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/992,334

FILING DATE: 17-DEC-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

US-08-992-334-3

APPLICATION NUMBER: US 08/302,752

FILING DATE: 24-DEC-1994

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/FR93/00248

FILING DATE: 12-MAR-1993

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 92/03034

FILING DATE: 13-MAR-1992

ATTORNEY/AGENT INFORMATION:

NAME: Prout, D. Bruce

REGISTRATION NUMBER: 20958

REFERENCE/DOCKET NUMBER: C93:31779

TELECOMMUNICATION INFORMATION:

TELEPHONE: (626) 795-9900

TELEFAX: (626) 577-8800

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 6722 base pairs

TYPE: nucleic acid

STRANDEDNESS: both

TOPOLOGY: circular

MOLECULE TYPE: DNA (genomic)

US-08-992-334-3

Query Match
Best Local Similarity 100.0%; Score 19; DB 2; Length 6722;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCCTGCAAAAGAGCAAG 1839
Db 4748 TCCTGCAAAAGAGCAAG 4766

RESULT 11

US-08-302-752-3
Sequence 3, Application US/08302752

Patent No. 6025190

GENERAL INFORMATION:

APPLICANT:

TITLE OF INVENTION: THERMOSENSIBLE PLASMID

NUMBER OF SEQUENCES: 3

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25 (PPO)

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/302,752

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: FR 9203034

FILING DATE: 13-MAR-1992

PRIOR APPLICATION DATA:

APPLICATION NUMBER: NO FR/93/00248

FILING DATE: 12-MAR-1993

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 6722 base pairs

TYPE: nucleic acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

US-08-302-752-3

Query Match
Best Local Similarity 100.0%; Score 19; DB 3; Length 6722;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCCTGCAAAAGAGCAAG 1839
Db 4748 TCCTGCAAAAGAGCAAG 4766

```
RESULT 12
US-08-778-494B-112
; Sequence 112, Application US/08778494B
; Patent No. 5962272
; GENERAL INFORMATION:
; APPLICANT: Chenchik, Alex
; APPLICANT: Zhu, York
; APPLICANT: Diachenko, Luda
; APPLICANT: Siebert, Paul
; TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR FULL-LENGTH CDNA
; TITLE OF INVENTION: CLONING
; NUMBER OF SEQUENCES: 114
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
; STREET: 2421 N.W. 41st Street, Suite A-1
; CITY: Gainesville
; STATE: Florida
; COUNTRY: USA
; ZIP: 32606
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION NUMBER: US/08/778,494B
; FILING DATE: 03-JAN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/582,562
; FILING DATE: 03-JAN-1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Pace, Doran R.
; REGISTRATION NUMBER: 38,261
; REFERENCE/DOCKET NUMBER: CL-7C1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (352) 375-8100
; TELEFAX: (352) 372-5800
; INFORMATION FOR SEQ ID NO: 112:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 23 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (synthetic)
US-08-778-494B-112

Query Match          0.9%; Score 18; DB 2; Length 23;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      21 CGGAAGCAGCAGAGG 38
        |||||||
Db      6 CGGAAGCAGCAGAGG 23

RESULT 13
US-09-045-284A-8/C
; Sequence 8, Application US/09045284A
; Patent No. 6265192
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerlich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107U51
; CURRENT APPLICATION NUMBER: US/09/045,284A
; CURRENT FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 47
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-045-284A-8

Query Match          0.9%; Score 18; DB 3; Length 47;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1248 ATCTTCGTCTACTCGA 1265
        |||||||
Db      47 ATCTTCGTCTACTCGA 30

RESULT 14
US-09-190-911-8/C
; Sequence 8, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerlich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 8
; LENGTH: 47
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-8

Query Match          0.9%; Score 18; DB 4; Length 47;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      1248 ATCTTCGTCTACTCGA 1265
        |||||||
Db      47 ATCTTCGTCTACTCGA 30

RESULT 15
US-09-439-313-320
; Sequence 320, Application US/09439313
; Patent No. 6329505
; GENERAL INFORMATION:
; APPLICANT: Xu, Jianshun
; APPLICANT: Dillon, Davin C.
; APPLICANT: Mitcham, Jennifer L.
; APPLICANT: Harlocker, Susan Louise
; APPLICANT: Jiang Yugu
; APPLICANT: Reed, Steven G.
; APPLICANT: Kalos, Michael
; APPLICANT: Fanger, Gary
; APPLICANT: Rietter, Mark
; APPLICANT: Day, Craig
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
; TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
; FILE REFERENCE: 210121.427C9
; CURRENT APPLICATION NUMBER: US/09/439,313
; CURRENT FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 575
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 320
; LENGTH: 150
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-439-313-320
```

Query Match 0.98; Score 18; DB 4; Length 150;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Caps 0;

OY 1 GAATCCATTGTTGGG 18
 ||||||||||||
 Db 31 GAATCCATTGTTGGG 48

Search completed: August 15, 2003, 00:46:53
 Job time : 143 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 20:11:42 ; Search time 7280 Seconds

(without alignments)
11480.545 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattccattgtgttgta.....tgggattcctaaaaaaa 2043

Scoring table: OLIGO_NMC

Gapop 60.0 , Gapext 60.0

Searched: 2888711 seqs, 2045481386 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5682662

Post-processing: Listing first 45 summaries

Database : GenEmbl:

1: gb_da:*
2: gb_htg:*
3: gb_in:*
4: gb_ov:*
5: gb_ov:*
6: gb_ov:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vl:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_mu:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vl:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rnd:*
36: em_htg_man:*
37: em_htg_vrl:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	1722	84.3	1992	9	AF280088	AF280088 Homo sapi
2	1653	80.9	2037	9	BC035282	BC035282 Homo sapi
3	1598	78.2	2011	9	AK026635	AK026635 Homo sapi
4	1597	78.2	183228	9	AC010547	AC010547 Homo sapi
5	1597	78.2	206943	2	AC138848	AC138848 Homo sapi
6	1547	75.7	1979	6	BD127258	BD127258 Homo sapi
7	1547	75.7	1979	6	AK074746	AK074746 Homo sapi
8	1547	75.7	2032	6	AR203335	AR203335 Sequence
9	1547	75.7	2032	6	BD134772	BD134772 Glycosyls
10	1547	75.7	2032	9	AF131235	AF131235 Homo sapi
11	1262	61.8	1333	9	AF149783	AF149783 Homo sapi
12	1153	56.4	2065	6	AR310485	AR310485 Sequence
13	695	34.0	877	6	BD124832	BD124832 Primer fo
14	695	34.0	877	6	BD126950	BD126950 Primer fo
15	403	19.7	517	6	AX381256	AX381256 Sequence
16	114	5.6	389	6	AX381326	AX381326 Sequence
17	105	5.1	548	6	BD125617	BD125617 Primer fo
18	59	2.9	1462	9	AF176839	AF176839 Homo sapi
19	59	2.9	1647	6	AX327330	AX327330 Sequence
20	59	2.9	2170	9	AF176838	AF176838 Homo sapi
21	59	2.9	2544	9	AF219990	AF219990 Homo sapi
22	59	2.9	3278	9	AF246718	AF246718 Homo sapi
23	59	2.9	3786	9	AF280086	AF280086 Homo sapi
24	59	2.9	71503	9	AF219991	AF219991 Homo sapi
25	59	2.9	157358	9	AC009163	AC009163 Homo sapi
26	59	2.9	194832	9	AC025287	AC025287 Homo sapi
27	59	2.9	208185	2	AC009105	AC009105 Homo sapi
28	37	1.8	37	6	AR203340	AR203340 Sequence
29	37	1.8	37	6	BD134774	BD134774 Glycosyls
30	30	1.5	30	6	BD134553	BD134553 Method fo
31	31	1.3	1926	6	BD134773	BD134773 Glycosyls
32	27	1.3	1926	10	AF131236	AF131236 Mus muscu
33	27	1.3	2201	10	AF109155	AF109155 Mus muscu
34	27	1.3	232745	2	AC130174	AC130174 Rattus no
35	27	1.3	262679	2	AC110314	AC110314 Rattus no
36	24	1.2	228569	2	AC121740	AC121740 Rattus no
37	24	1.2	235173	2	AC133449	AC133449 Rattus no
38	24	1.2	240901	2	AC096328	AC096328 Rattus no
39	24	1.2	242253	2	AC095651	AC095651 Rattus no
40	24	1.2	246877	2	AC137365	AC137365 Rattus no
41	24	1.2	258097	2	AC129764	AC129764 Rattus no
42	24	1.2	271230	2	AC094919	AC094919 Rattus no
43	23	1.1	1740	10	AF176841	AF176841 Mus muscu
44	23	1.1	1989	10	AF176840	AF176840 Mus muscu
45	23	1.1	110000	2	AC099262_1	Continuation (2 of

ALIGNMENTS

RESULT 1
AF280088
LOCUS
DEFINITION Homo sapiens l-selectin ligand sulfotransferase GST-3 mRNA,
complete cds.
ACCESSION AF280088
VERSION AF280088.1 GI:12060807
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM
REFERENCE
AUTHORS
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 1992)
Hemmerich,S., Lee,J.K., Bhakta,S., Bistrup,A., Ruddle,N.R. and
Rosen,S.D.

TITLE Chromosomal localization and genomic organization for the
galactose/N-acetylglucosamine/N-acetylglucosamine
6-O-sulfotransferase gene family
JOURNAL Glycobiology 11 (1), 75-87 (2001)
MEDLINE 21096027
PUBMED 11815564
REFERENCE 2 (bases 1 to 1992)
AUTHORS Hemmerlich, S., Bhakta, S., Lee, J.-K., Bistrup, A., Ruddle, N.R. and
Rosen, S.D.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience,
3401 Hillview Ave., Palo Alto, CA 94304, USA
FEATURES
source
1..1992
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/chromosome="16"
/map="16q23.1-q23.2"
/cell_type="high endothelial"
/tissue_type="tonsil"
1..1992
/gene="GST3"
122..1282
/note="similar to Homo sapiens sequence encoded by Genbank
Accession Number AF131235"
/product="L-selectin ligand sulfotransferase GST-3"
/protein_id="AA048246.1"
/db_xref="GI:12060808"
/translation="MLLPKRMILFLVDSQALILAFPHSHNISLSMKQPERH
VLVLSMRSGSFVGOLEPQHPDVFILMEPAHVMWTEKOSTAMMLHNAVRLIRAVE
LCDMSVFDAYMEPRGPRROSLPQWENSRLCSAPACDILPDEILIPRAKRLISQOP
FEYVAKRASHVYLKEVFENLQSLYLDKPSILNLIHLVDRPAVFRSRETK
GOLMIDSRIVMGHOEKLKEDOPYVMOVICOSOLEIKTISLPKAOEYELVRY
EDLAPAPVOTSRMTEFVLEFLPHLOFVHNITRGKMGDIAFPHNADALNVQAM
RMSLEYKYSRQKAGDMANLIGTRHRSBEDQRNLIDLILSTVTPQIHT"

BASE COUNT 482 a 540 c 489 g 481 t
ORIGIN

Query Match 84.3%; Score 1722; DB 9; Length 1992;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

7 CATTGTGTTGGGTACGGAAGACGACGAGAGGGTAGAGGAAAGCCGATGCCCCGCTA 66
1 CATTGTGTTGGGTACGGAAGACGACGAGAGGGTAGAGGAAAGCCGATGCCCCGCTA 60

67 GGAGTGAAGCTCTCAAAAGCAGCAGGGAAGCCCAAGCTTCCACTTCAGCAC 126
61 GCAGTGAGCTCTCAAAAGCAGCAGGGAAGCCCAAGCTTCCACTTCAGCAC 120

127 AATGCTACTGCTTAAATAAGCTCTGCTGTTTGGTTTCCAGATGGCATCTT 186
121 AATGCTACTGCTTAAATAAGCTCTGCTGTTTGGTTTCCAGATGGCATCTT 180

187 GGCCTATTCTCCACATGTACAGCCACAACATCAGCTCCCTGTATGAAGCAGACC 246
181 GGCCTATTCTCCACATGTACAGCCACAACATCAGCTCCCTGTATGAAGCAGACC 240

247 CGAGGGCATGACGCTGCTGCTCTTCCGCGGCTTGGCTCTCTTTTGTGGGCA 306
241 CGAGGGCATGACGCTGCTGCTCTTCCGCGGCTTGGCTCTCTTTTGTGGGCA 300

307 GCTTTTGGGAGACCCAGATGTTTCTTACTGATGAGGAGCCGCTGGCAGTGGAT 366
301 GCTTTTGGGAGACCCAGATGTTTCTTACTGATGAGGAGCCGCTGGCAGTGGAT 360

367 GACCTTCAAGCAGACCGCTGGATGCTGCACATGGCTGCGGATCTATACGGAC 426
361 GACCTTCAAGCAGACCGCTGGATGCTGCACATGGCTGCGGATCTATACGGAC 420

427 CGTCTTCTTGCGCAGCATGAGCGCTTTTGATGCTTACATGAACTGTGCCCCGAGACA 486
421 CGTCTTCTTGCGCAGCATGAGCGCTTTTGATGCTTACATGAACTGTGCCCCGAGACA 480

487 GTCCAGCCCTTTTCAGTGGGAGAACAGCCGGGCCCTGTGTGTGCAACCTGTGCAT 546
481 GTCCAGCCCTTTTCAGTGGGAGAACAGCCGGGCCCTGTGTGTGCAACCTGTGCAT 540

547 CATCCCAAGAAGTGAATCATCCCGGGGCTCACTGACAGCTCTGTGAGTCAACAGCC 606
541 CATCCCAAGAAGTGAATCATCCCGGGGCTCACTGACAGCTCTGTGAGTCAACAGCC 600

607 CTTTGAAGTGTGGAGAAAGCCCTGCCCTCTTACAGCCACGTGTGTCTTAAGAGGTGG 666
601 CTTTGAAGTGTGGAGAAAGCCCTGCCCTCTTACAGCCACGTGTGTCTTAAGAGGTGG 660

667 CTTCTTCAACCTGACAGTCCCTGTACCCGCTGTGTAAGAGCCCTCCCTCACTGATAT 726
661 CTTCTTCAACCTGACAGTCCCTGTACCCGCTGTGTAAGAGCCCTCCCTCACTGATAT 720

727 CGTGCACCTGTGTCGGGAGACCCCGGGCCGTGTTCGCTCCGAGAAAGCAAAAGGAGA 786
721 CGTGCACCTGTGTCGGGAGACCCCGGGCCGTGTTCGCTCCGAGAAAGCAAAAGGAGA 780

787 TCTCATGATTGACAGTCGCAATTGTGATGGGGCAGCATGAGCAAAACTCAAGAGAGA 846
781 TCTCATGATTGACAGTCGCAATTGTGATGGGGCAGCATGAGCAAAACTCAAGAGAGA 840

847 CCAACCCCTCTATGATGATGATCAGTCACTGCGCAAGCCAGCTGAGATCTACAAGCAT 906
841 CCAACCCCTCTATGATGATGATCAGTCACTGCGCAAGCCAGCTGAGATCTACAAGCAT 900

907 CGAGTCTTGCCCAAGAGCCCTCAGAGAACGTAACCTGCTTGGCGTATGAGACCTGGC 966
901 CGAGTCTTGCCCAAGAGCCCTCAGAGAACGTAACCTGCTTGGCGTATGAGACCTGGC 960

967 TCGAGCCCTGTGGGCCCAAGCTTCCCGAATGTATGATTCGTGGGATTTGGAATTTGCGC 1026
961 TCGAGCCCTGTGGGCCCAAGCTTCCCGAATGTATGATTCGTGGGATTTGGAATTTGCGC 1020

1027 CCATCTTACAGCTGGGTGATATACATCAACCCGAGGCAAGGATGAGGACACGCTTT 1086
1021 CCATCTTACAGCTGGGTGATATACATCAACCCGAGGCAAGGATGAGGACACGCTTT 1080

1087 CCACACAAATGCCAGGATGCCCTTAATGTCTCCAGAGCTTGGCGTGTCTTCCCTA 1146
1081 CCACACAAATGCCAGGATGCCCTTAATGTCTCCAGAGCTTGGCGTGTCTTCCCTA 1140

1147 TGAAGAGTTTCTCCAGCTTCAAGAAAGCCGTGTGGCATGCAATGATTTGCTGGCTACCG 1206
1141 TGAAGAGTTTCTCCAGCTTCAAGAAAGCCGTGTGGCATGCAATGATTTGCTGGCTACCG 1200

1207 CCAGCTCAGATCTGAACAAGACAGAGAAACTGTTGCGATCTTCTCTTCACTGGAC 1266
1201 CCAGCTCAGATCTGAACAAGACAGAGAAACTGTTGCGATCTTCTCTTCACTGGAC 1260

1267 TGTCCCTGAGCAATCACTAAGAGGTTGAGAGGCTTGTGTCACCTGTGTGACGC 1326
1261 TGTCCCTGAGCAATCACTAAGAGGTTGAGAGGCTTGTGTCACCTGTGTGACGC 1320

1327 TCAAGTCACTTCTCTGAATGCTTCTGAGACCTTGGCTTAATCTCTAGAGCTTAATG 1386
1321 TCAAGTCACTTCTCTGAATGCTTCTGAGACCTTGGCTTAATCTCTAGAGCTTAATG 1380

1387 TCTGTGGGATACACTGAGTGTGATGTGTCCACAGTGTGTCGAAGAGAAAGACTTT 1446
1381 TCTGTGGGATACACTGAGTGTGATGTGTCCACAGTGTGTCGAAGAGAAAGACTTT 1440

1447 GTGTCCATGCTGTGTCTTGAAGAACAGACTGGGGAACCTTATGTGAGCAGCAATCCAC 1506
1441 GTGTCCATGCTGTGTCTTGAAGAACAGACTGGGGAACCTTATGTGAGCAGCAATCCAC 1500

1507 CAGTGAACAGGATATGCTCTTCTTCTTCTGATCTTCTCTGCTGGGAGACTTCAG 1566

```

Db      1501 CAGTGAACAGGATATGCTCTCTCTTCTTCTTATCTCTGCTGCGGACAGCTCAG 1560
QY      1567 AGACTTTGTGGCTCGAGGCTCTATTAAGCAGACAGCTATCACTGGAATGATCAATA 1626
Db      1561 AGACTTTGTGGCTCGAGGCTCTATTAAGCAGACAGCTATCACTGGAATGATCAATA 1620
QY      1627 ACCTCCCTGTCACATCTTGTCCCAATGGGGAATGATCTTCAACCAAGAGCTCACAGC 1686
Db      1621 ACCTCCCTGTCACATCTTGTCCCAATGGGGAATGATCTTCAACCAAGAGCTCACAGC 1680
QY      1687 ATTTTCACAGAGATGCGAATTTCTGAGCCCTTGAGATTCCA 1728
Db      1681 ATTTTCACAGAGATGCGAATTTCTGAGCCCTTGAGATTCCA 1722

RESULT 2
BC035282 2037 bp mRNA linear PRI 23-SEP-2002
LOCUS     Homo sapiens, similar to carbohydrate (N-acetylglucosamine 6-O)
DEFINITION sulfotransferase 4, clone MGC:34456 IMAGE:5179826, mRNA, complete
cds
ACCESSION BC035282
VERSION    BC035282.1 GI:23273964
KEYWORDS   MGC.
SOURCE     Homo sapiens (human)
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2037)
AUTHORS   Strausberg, R.
TITLE     Direct Submission
JOURNAL    Submitted (31-JUL-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2550,
USA
REMARK     NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT    Contact: MGC help desk
Email: gclabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: http://www.hgsc.bcm.tmc.edu/cdna/
Contact: amgebcm.tmc.edu
Guaratine, P.H., Garcia, A.M., Lu, X., Huiyik, S.W., Hale, S.M.,
Toon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LNL at: http://image.lnl.gov
Series: IRK Plate: 50 Row: a Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 5031734.
Location/Qualifiers
1. 2037
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:34456 IMAGE:5179826"
/tissue_type="Brain, Lung, Testis, adult, pooled whole"
/clone_lib="NIH_MGC_115"
/lab_host="DH10B"
/note="Vector: PCMV-SPORT6"
126..1238
/codon_start=1
/product="Similar to carbohydrate (N-acetylglucosamine
6-O) sulfotransferase 4"
/protein_id="AAH35282.1"
/db_xref="GI:23273965"
/translation="MAIIALFPMYSHNISLSMKAQPERHVLVLSWRSGLSFWGQ

```

```

LFGHPDVFYIMEPAHVMNTFKOSTAMLMHVARDLIRAVFLCDMSVFDAYNEPGR
ROSLTFQWENSRLALCSAPACDIDIPDILIRARHRLCSOPEPVKACRSYSHVVL
KEYREFNQLSLYPLKNDPSLNHIVHVRPRPRAVRSREKTDLMDSRYMGOHQD
KLKREDPIYVMQYTCOSLEIKTYSLSRPAERLIVLEIDLAIPAQSRME
FVGLIEFLPILOTWHNTIRKGMGDHFNHARALVNSQAMRWSLPEYKVSRLQKAC
GDAMNLLGYRHVSEQDORNLIDLLSTWTVPEQIH"
BASE COUNT 507 a 546 c 481 g 503 t
ORIGIN
Query Match 80.9%; Score 1653; DB 9; Length 2037;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1653; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
1
111 TCTTCCACTTCAGCACAATGCTACTGCTAATAAATAAGAGCTCTGTTCTGGTTT 170
Db      1 GCGCATGGCCGGCTTACAGTAGAGCCCTCTCAAAAGCAGGAGGAGCCCAAGCAAG 60
QY      171 CCCAGATGGCCATTTGGCTCTATTCTTCCATGTACAGCCACACATCAGTCCCTGT 230
Db      121 CCCAGATGGCCATTTGGCTCTATTCTTCCATGTACAGCCACACATCAGTCCCTGT 180
QY      231 CTATGAAGGCACAGCCCGAGGAGCATGACGCTGTGTTCTGCTGGGCTGTGGCT 290
Db      181 CTATGAAGGCACAGCCCGAGGAGCATGACGCTGTGTTCTGCTGGGCTGTGGCT 240
QY      291 CTCTCTTTTGTGGGGACCTTTTGGGAGCACCAGATGTTTCTTACCTGATGAGCCG 350
Db      241 CTCTCTTTTGTGGGGACCTTTTGGGAGCACCAGATGTTTCTTACCTGATGAGCCG 300
QY      351 CTTGCGACGTGTGATATACCTTCAAGAGACAGCCGCTGGATCTGCATGAGCTGTC 410
Db      301 CTTGCGACGTGTGATATACCTTCAAGAGACAGCCGCTGGATCTGCATGAGCTGTC 360
QY      411 GGGATCTGATACGGGCGCTCTTCTTGCGCATGAGACGCTTGTGATGCTACATGGAAC 470
Db      361 GGGATCTGATACGGGCGCTCTTCTTGCGCATGAGACGCTTGTGATGCTACATGGAAC 420
QY      471 CTGTGCCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACAGCCGGCCCTGTGTCTG 530
Db      421 CTGTGCCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACAGCCGGCCCTGTGTCTG 480
QY      531 CACCTGCTGTGATATATCCACAAGATGAAATATCCCGGGCTCTACGCAAGCTCC 590
Db      481 CACCTGCTGTGATATATCCACAAGATGAAATATCCCGGGCTCTACGCAAGCTCC 540
QY      591 TGTGCACTCAACAGCCCTTGTGAGGTGGAGAGAGCCGCTGCTACAGCAGCTGG 650
Db      541 TGTGCACTCAACAGCCCTTGTGAGGTGGAGAGAGCCGCTGCTACAGCAGCTGG 600
QY      651 TGTGCAAGAGAGTGGCTTCTTCAACTGCAGTCCCTCTACCGCTCTGAAGACCTT 710
Db      601 TGTGCAAGAGAGTGGCTTCTTCAACTGCAGTCCCTCTACCGCTCTGAAGACCTT 660
QY      711 CCTTCACCTGCAATATGCTGACCTGCTCGGAGACCCCGGGCGCTGCTCCGGAG 770
Db      661 CCTTCACCTGCAATATGCTGACCTGCTCGGAGACCCCGGGCGCTGCTCCGGAG 720
QY      771 AACGCAAAAGGAGATCTCATGTATGACAGTGCATTTGATGGGAGCAGATGAGCAAA 830
Db      721 AACGCAAAAGGAGATCTCATGTATGACAGTGCATTTGATGGGAGCAGATGAGCAAA 780
QY      831 AACTCAAGAAAGAGACCAACCTTACTATGTATGACAGTCAATCTGCCAAAGCCAGCTGG 890
Db      781 AACTCAAGAAAGAGACCAACCTTACTATGTATGACAGTCAATCTGCCAAAGCCAGCTGG 840
QY      891 AGATCTCAAGACCATCATCAGTCTTGGCCAAAGGCCGCGAGGAAGCTTACTGCTTGGC 950
Db      841 AGATCTCAAGACCATCATCAGTCTTGGCCAAAGGCCGCGAGGAAGCTTACTGCTTGGC 900

```



```

OY 951 GGTATGAGACCTGGCTGAGCCCTGTGTGCCCCAGACTTCCGATGTATGAAATTCGTGG 1010
Db 901 GGTATGAGACCTGGCTGAGCCCTGTGTGCCCCAGACTTCCGATGTATGAAATTCGTGG 960
OY 1011 GATTGGAATTCCTGGCCCATCTTCAGACCTGGGTGATATCAACCCAGGCAAGGGCA 1070
Db 961 GATTGGAATTCCTGGCCCATCTTCAGACCTGGGTGATATCAACCCAGGCAAGGGCA 1020
OY 1071 TGGGTGACACAGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGC 1130
Db 1021 TGGGTGACACAGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGC 1080
OY 1131 GGTGCTCTTTGGCCCATGAAAAGTTTCTGACATTCAGAAAAGCTGTGGCATCATATA 1190
Db 1081 GGTGCTCTTTGGCCCATGAAAAGTTTCTGACATTCAGAAAAGCTGTGGCATCATATA 1140
OY 1191 ATTGCTGGGCTACCGCCACGTCAGATCTGAACAAGAAAGAAACCTGTGCTGGATC 1250
Db 1141 ATTGCTGGGCTACCGCCACGTCAGATCTGAACAAGAAAGAAACCTGTGCTGGATC 1200
OY 1251 TTCTGTACACAGACCTGCTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAGCTTGTCTG 1310
Db 1201 TTCTGTACACAGACCTGCTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAGCTTGTCTG 1260
OY 1311 CCACCTGGTGTAGGCTCAGTCACTTCTCTGAATGCTTCTGAGCCTTCCATCATCTCT 1370
Db 1261 CCACCTGGTGTAGGCTCAGTCACTTCTCTGAATGCTTCTGAGCCTTCCATCATCTCT 1320
OY 1371 GAGCCTTAACATAGTGTGTGGTATCACTAGTGTGAGTGTGTGTCACACGCTGC 1430
Db 1321 GAGCCTTAACATAGTGTGTGGTATCACTAGTGTGAGTGTGTGTCACACGCTGC 1380
OY 1431 AAGCAGAAAGACTTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGAACTTATCT 1490
Db 1381 AAGCAGAAAGACTTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGAACTTATCT 1440
OY 1491 GAGCAGACATCCCAACAGATGAAGGATTTGCTCTTCTTCTTGTGATCTTCTG 1550
Db 1441 GAGCAGACATCCCAACAGATGAAGGATTTGCTCTTCTTCTTGTGATCTTCTG 1500
OY 1551 TCTGGGAGACTTTCAGAGACTTGTGGCTTGAGGCGCTTAAAGCAGACAGATATCAG 1610
Db 1501 TCTGGGAGACTTTCAGAGACTTGTGGCTTGAGGCGCTTAAAGCAGACAGATATCAG 1560
OY 1611 TGGAAATGATCCATAAACCCTCCCTGTCCACATCTTGCCCAATGGGGAATGATCTTTCAC 1670
Db 1561 TGGAAATGATCCATAAACCCTCCCTGTCCACATCTTGCCCAATGGGGAATGATCTTTCAC 1620
OY 1671 CAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
Db 1621 CAAAGAGCTCACAGCATTTTCCACAGAGATGC 1653

```

```

RESULT 3
ACCESSION AK026635 2011 bp mRNA linear PRI 29-SEP-2000
VERSION AK026635.1 GI:10439531
LOCUS Homo sapiens cDNA: FLJ22982 fis, clone KAT11454, highly similar to
DEFINITION AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase
mRNA.
SOURCE Homo sapiens
ORGANISM Homo sapiens (human)
REFERENCE 1 (sites)
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
Matsunabe, K., Kumagai, A., Itakura, S., Yamazaki, M., Tashiro, H.,
Ota, T., Suzuki, Y., Obayashi, M., Nishi, T., Shibahara, T., Tanaka, T.,
Nakamura, Y., Isogai, T. and Sugano, S.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 2011)

```

```

AUTHORS Sugano, S., Suzuki, Y., Ota, T., Obayashi, M., Nishi, T., Isogai, T.,
Shibahara, T., Tanaka, T. and Nakamura, Y.
TITLE Direct Submission
JOURNAL Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail: cdna@ims.u-tokyo.ac.jp, Tel: 81-3-5449-5286,
Fax: 81-3-5449-5416)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan; cDNA full insert
sequencing: Research Association for Biotechnology; cDNA library
construction, 5'-3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source location/Qualifiers
1..2011
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="KAT11454"
/cell_line="KATO III"
/cell_type="signed-ring cell carcinoma"
/clone_1fb="KAT"
/clone_2fb="KAT"
/notes="cloning vector pME18SFL3"
1..2011
/notes="highly similar to AF131235 Homo sapiens
N-acetylglucosamine 6-O-sulfotransferase mRNA"
BASE COUNT 482 a 541 c 475 g 513 t
ORIGIN
Query Match 78.2%; Score 1598; DB 9; Length 2011;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1598; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 106 CAAGTCTTCACCTTACAGCAATGCTACTGCTTAAAGAAAGTCTGCTGTTCT 165
Db 50 CAAGTCTTCACCTTACAGCAATGCTACTGCTTAAAGAAAGTCTGCTGTTCT 109
OY 166 GGTTCGCCAGATGGCCATCTTGCTATCTTCCACATGATACAGCACAATCAGCTC 225
Db 110 GGTTCGCCAGATGGCCATCTTGCTATCTTCCACATGATACAGCACAATCAGCTC 169
OY 226 CCTGTCTAAGAGGACAGCCCGAGCGCATGCACTGCTGTCTCTCTCTGCGCTC 285
Db 170 CCTGTCTAAGAGGACAGCCCGAGCGCATGCACTGCTGTCTCTCTCTGCGCTC 229
OY 286 TGGCTCTCTTTTGTGGGAGCAGTTTGTGGGAGCAGCCAGATGTTTCTACCTGATGA 345
Db 230 TGGCTCTCTTTTGTGGGAGCAGTTTGTGGGAGCAGCCAGATGTTTCTACCTGATGA 289
OY 346 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGACAGCCGCTGATGTCACATGCG 405
Db 290 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGACAGCCGCTGATGTCACATGCG 349
OY 406 TGTGGGAGTGTGATACGCGCGCTCTTCTGTGCGACATGAGCGCTTGTGATGCTACT 465
Db 350 TGTGGGAGTGTGATACGCGCGCTCTTCTGTGCGACATGAGCGCTTGTGATGCTACT 409
OY 466 GGAACCTGTGCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACAGCGCGGCTCGTG 525
Db 410 GGAACCTGTGCCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAACAGCGCGGCTCGTG 469
OY 526 TTCTGACCTGCTGTGATCATCCACAAATGCAATCAATCCCGGGGCTCACTGACG 585
Db 470 TTCTGACCTGCTGTGATCATCCACAAATGCAATCAATCCCGGGGCTCACTGACG 529
OY 586 GCTCTGTGACATCAACAGCCCTTGTAGGTGTGGAAGAGCGCTGCTCTACAGCA 645
Db 530 GCTCTGTGACATCAACAGCCCTTGTAGGTGTGGAAGAGCGCTGCTCTACAGCA 589
OY 646 GGTGGTCAAGAGGAGTGGCTTCAACTGAGTCCCTACCGGCTGCTGAAGA 705

```

Db 590 CGTGTGCTCAGAGAGTGGCTTCTTCAACCTGCACTGCCCTCTACCCGCTGCTGANAAGA 649
 QY 706 CCCCTCCCTCAACCTGCATATCGNACGCTGTCGGGAGACCCCGGCGGCTGTTCCGTTTC 765
 Db 650 CCCCTCCCTCAACCTGCATATCGTGCACCTGTCGGGAGACCCCGGCGGCTGTTCCGTTTC 709
 QY 766 CCGAGAACGCAAAAGGAGATCTCATGATGTGACAGTGCATGTTGATGGGGGAGCATGA 825
 Db 710 CCGAGAACGCAAAAGGAGATCTCATGATGTGACAGTGCATGTTGATGGGGGAGCATGA 769
 QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTATGATGAGTCTATCTGCGCAACCA 885
 Db 770 GCAAAAACCTCAAGAGAGAGACCAACCTTACTATGTATGATGAGTCTATCTGCGCAACCA 829
 QY 886 GCTGAGATCTACAAAGACATCCAGTCTTCCCAAGGCGCTGAGAGAGCGCTACTGCT 945
 Db 830 GCTGAGATCTACAAAGACATCCAGTCTTCCCAAGGCGCTGAGAGAGCGCTACTGCT 889
 QY 946 TGTGCGCTATGAGAGACCTGCTGAGCCCTGTGCGCCAGACTTCCGAAATGATGAAT 1005
 Db 890 TGTGCGCTATGAGAGACCTGCTGAGCCCTGTGCGCCAGACTTCCGAAATGATGAAT 949
 QY 1006 CGTGGGATTTGGAATCTTGGCCCATCTTCAAGACCTGGGTGATTAACATCAACCCGAGCA 1065
 Db 950 CGTGGGATTTGGAATCTTGGCCCATCTTCAAGACCTGGGTGATTAACATCAACCCGAGCA 1009
 QY 1066 GGGCATGGTGACCAACCTTTCACACAAATGCGAGGATGCGCTTATGTCTCCAGGC 1125
 Db 1010 GGGCATGGTGACCAACCTTTCACACAAATGCGAGGATGCGCTTATGTCTCCAGGC 1069
 QY 1126 TTGGCGCTGCTCTTGGCCCATCTTCAAGACCTGGGTGATTAACATCAACCCGAGCA 1185
 Db 1070 TTGGCGCTGCTCTTGGCCCATCTTCAAGACCTGGGTGATTAACATCAACCCGAGCA 1129
 QY 1186 CATGAATTTGCTGGGCTACCGCCAGCTGATGTAACAGAGAAACCTGTTGCT 1245
 Db 1130 CATGAATTTGCTGGGCTACCGCCAGCTGATGTAACAGAGAAACCTGTTGCT 1189
 QY 1246 GGAATCTTCTGCTACCTGAGTGTCTCCAGCAATGCGAGGATGCGCTTATGTCTCCAGGC 1305
 Db 1190 GGAATCTTCTGCTACCTGAGTGTCTCCAGCAATGCGAGGATGCGCTTATGTCTCCAGGC 1249
 QY 1306 TGTGCGCACTGGTGTGAGCCTCAGTCACTTCTGTAATGCTTCTGAGCCTTCTTACA 1365
 Db 1250 TGTGCGCACTGGTGTGAGCCTCAGTCACTTCTGTAATGCTTCTGAGCCTTCTTACA 1309
 QY 1366 TGTGAGCCTTACTACATGCTGCTGTTGATGATCACTGATGATGTTGTTCCACAGC 1425
 Db 1310 TGTGAGCCTTACTACATGCTGCTGTTGATGATCACTGATGATGTTGTTCCACAGC 1369
 QY 1426 TGTCAAGCAGAGAGACTTGTGTCATGCTTGTCTAGAAACAGACTGGGGAACT 1485
 Db 1370 TGTCAAGCAGAGAGACTTGTGTCATGCTTGTCTAGAAACAGACTGGGGAACT 1429
 QY 1486 TATGTGAGCAGACATCCCAAGTGAAGAGGGTATGCTTCTTCTTCTTGTGATCT 1545
 Db 1430 TATGTGAGCAGACATCCCAAGTGAAGAGGGTATGCTTCTTCTTCTTGTGATCT 1489
 QY 1546 TCTCTGTGGGAGAGACTTCAAGAGCTTGTGGCTGAGAGCCTTATTAAGCAGCAGACT 1605
 Db 1490 TCTCTGTGGGAGAGACTTCAAGAGCTTGTGGCTGAGAGCCTTATTAAGCAGCAGACT 1549
 QY 1606 ATCAGTGAATTTGATCATTAACCTGCTGTCATCTGCGCAATGCGGGAATGATCT 1665
 Db 1550 ATCAGTGAATTTGATCATTAACCTGCTGTCATCTGCGCAATGCGGGAATGATCT 1609
 QY 1666 TTCACCAAGAGCTCACAGACTTTCACAGAGATGCT 1703
 Db 1610 TTCACCAAGAGCTCACAGACTTTCACAGAGATGCT 1647

RESULT 4
 AC010547

LOCUS AC010547 183228 bp DNA linear PRI 29-SEP-2001
 DEFINITION Homo sapiens chromosome 16 clone RP11-510M2, complete sequence.
 AC010547
 VERSION AC010547.9 GI:15808510
 KEYWORDS HTG.
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE Sequencing of Human Chromosome 16
 JOURNAL Unpublished
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE 2 (bases 1 to 183228)
 JOURNAL Direct Submission
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE 3 (bases 1 to 183228)
 JOURNAL Direct Submission
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE Submitted (29-SEP-2001) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 On Sep 29, 2001 this sequence version replaced g1:14589436.
 COMMENT
 JOURNAL
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 GenBank flat file format but are available as part
 of this entry's ASN.1 file.

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 GenBank flat file format but are available as part
 of this entry's ASN.1 file.

 FEATURES
 source
 1. 183228
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-510M2"
 BASE COUNT 51901 a 42098 c 40026 g 49203 t
 ORIGIN
 Query Match 78.2%; Score 1597; DB 9; Length 183228;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 107 AAGCTCTTCACCTTCAGCAGATGCTACTGCTTAAAAAATGAAGCTCTCTGTTCTG 166
 Db 142584 AAGCTCTTCACCTTCAGCAGATGCTACTGCTTAAAAAATGAAGCTCTCTGTTCTG 142643
 QY 167 GTTCCAGAGGGCAGCTTGGCTATCTTCCACATGTCACGCAACATCAGCTCC 226
 Db 142644 GTTCCAGAGGGCAGCTTGGCTATCTTCCACATGTCACGCAACATCAGCTCC 142703
 QY 227 CTGCTATGAAGGACAGCCCGAGCGATGACAGCTGTGTTCTCTTCCGCGCTCT 286
 Db 142704 CTGCTATGAAGGACAGCCCGAGCGATGACAGCTGTGTTCTCTTCCGCGCTCT 142763
 QY 287 GGCCTCTCTTTTGTGGGGAGCTTTTGGGAGACCCAGATGTTTCTACCTGATGGAG 346
 Db 142764 GGCCTCTCTTTTGTGGGGAGCTTTTGGGAGACCCAGATGTTTCTACCTGATGGAG 142823

```

OY 347 CCCGCTGGCAGGTGTGATGACCTTCAAGACAGACCGCCTGATGCTGCACATGGCT 406
    |||||
Db 142824 CCCGCTGGCAGGTGTGATGACCTTCAAGACAGACCGCCTGATGCTGCACATGGCT 142883
OY 407 GTGCGGGATCTGATACAGGGCGCTCTTCTGTGACATGAGCGCTTTTATGCTTACATG 466
    |||||
Db 142884 GTGCGGGATCTGATACAGGGCGCTCTTCTGTGACATGAGCGCTTTTATGCTTACATG 142943
OY 467 GAACCTGTGCCCCGGAGACATCCAGCCCTTTCAGTGGGAGAACGCGGGCGCTGTGT 526
    |||||
Db 142944 GAACCTGTGCCCCGGAGACATCCAGCCCTTTCAGTGGGAGAACGCGGGCGCTGTGT 143003
OY 527 TCTGCACTGCTGTGATCATCCACAGATGAATCATCCCCGGGCTCACTGTCAG 586
    |||||
Db 143004 TCTGCACTGCTGTGATCATCCACAGATGAATCATCCCCGGGCTCACTGTCAG 143063
Y 587 CTCCTGTGAGTCAACAGCCCTTTGAGGTGTGGAGAGAGCGCTGCCCTCTTACAGCCAC 646
    |||||
Db 143064 CTCCTGTGAGTCAACAGCCCTTTGAGGTGTGGAGAGAGCGCTGCCCTCTTACAGCCAC 143123
OY 647 GTGTGCTCAAGAGAGTGGCTTCTTCACTGACATGCTCCCTGTACCCGCTGGAAGAC 706
    |||||
Db 143124 GTGTGCTCAAGAGAGTGGCTTCTTCACTGACATGCTCCCTGTACCCGCTGGAAGAC 143183
OY 707 CCTCTCCTTCACTGATATGCTGCACTGTGCGGAGCCCGGGCGCTGTCCGCTTCC 766
    |||||
Db 143184 CCTCTCCTTCACTGATATGCTGCACTGTGCGGAGCCCGGGCGCTGTCCGCTTCC 143243
OY 767 CGAGAACGCAAAAGGAGATCTCATGATGTGACAGTGTGATGTGGGGAGCATGTAG 826
    |||||
Db 143244 CGAGAACGCAAAAGGAGATCTCATGATGTGACAGTGTGATGTGGGGAGCATGTAG 143303
OY 827 CAAAACTCAAAAGAGAGACCAACCTCTATGTGATGATGATGATGATGATGATGATGATG 886
    |||||
Db 143304 CAAAACTCAAAAGAGAGACCAACCTCTATGTGATGATGATGATGATGATGATGATGATG 143363
OY 887 CTGAGATCTCAAAAGAGATCTCAAGTCTTGGCCAGAGCGCTTCAAGATGATGATGATG 946
    |||||
Db 143364 CTGAGATCTCAAAAGAGATCTCAAGTCTTGGCCAGAGCGCTTCAAGATGATGATGATG 143423
OY 947 GTGCGCTATGAGAGCTGTGCTGAGCGCCCTGTGCGCCAGATCTCCGATGATGATGATG 1006
    |||||
Db 143424 GTGCGCTATGAGAGCTGTGCTGAGCGCCCTGTGCGCCAGATCTCCGATGATGATGATG 143483
OY 1007 GTGAGATGGAATCTTGGCCCATCTTCAAGCTGTGGTGCATTAACATCAACCGAGCAAG 1066
    |||||
Db 143484 GTGAGATGGAATCTTGGCCCATCTTCAAGCTGTGGTGCATTAACATCAACCGAGCAAG 143543
OY 1067 GGCAATGGGTGACAGGCTTTCACACAAATGCGAGGATGCCCTTAATGCTTCCAGGCT 1126
    |||||
Db 143544 GGCAATGGGTGACAGGCTTTCACACAAATGCGAGGATGCCCTTAATGCTTCCAGGCT 143603
OY 1127 TGGCGCTGTCTTTCCTCATGAAAGGTTTCTGCACTTCAAGAAAGCTGTGGCGATGCG 1186
    |||||
Db 143604 TGGCGCTGTCTTTCCTCATGAAAGGTTTCTGCACTTCAAGAAAGCTGTGGCGATGCG 143663
OY 1187 ATGAATTTGCTGGGCTACCGCACGTCAGATGTGAACAAGAACAGAAACCTGTGTCG 1246
    |||||
Db 143664 ATGAATTTGCTGGGCTACCGCACGTCAGATGTGAACAAGAACAGAAACCTGTGTCG 143723
OY 1247 GATCTTGTCTACTGAGCTTCCCTTGAGCAATTCATTAAGAGGTTGAAGAGCTTT 1306
    |||||
Db 143724 GATCTTGTCTACTGAGCTTCCCTTGAGCAATTCATTAAGAGGTTGAAGAGCTTT 143783
OY 1307 GGTGCACTGTGTGCTGAGCTTCTGCACTTCTGCAATGCTTCTGAGGCTTGCTCAAT 1366
    |||||
Db 143784 GGTGCACTGTGTGCTGAGCTTCTGCACTTCTGCAATGCTTCTGAGGCTTGCTCAAT 143843
OY 1367 CTCTAGACCTTAACATGATGTGTGGGTATCAACACTGATGTGATGTGTGTCACACGT 1426
    |||||
Db 143844 CTCTAGACCTTAACATGATGTGTGGGTATCAACACTGATGTGATGTGTGTCACACGT 143903

```

```

OY 1427 GCTCAGCAGAAAGAGCTTTTGTCTCCATGCTTGTGTCTAGAAACAGACTGGGAACTT 1486
    |||||
Db 143904 GCTCAGCAGAAAGAGCTTTTGTCTCCATGCTTGTGTCTAGAAACAGACTGGGAACTT 143963
OY 1487 ATGTGAGCAGACATCCACACAGTGAACAGGATATGCTCTTCTTCTTGTATCTT 1546
    |||||
Db 143964 ATGTGAGCAGACATCCACACAGTGAACAGGATATGCTCTTCTTCTTGTATCTT 144023
OY 1547 CCTGTGCGCAGACTTCAAGAGCTTGTGGCTGAGGCGTATTAAGCAGCAGACATG 1606
    |||||
Db 144024 CCTGTGCGCAGACTTCAAGAGCTTGTGGCTGAGGCGTATTAAGCAGCAGACATG 144083
OY 1607 TCAGTGAATGATATCAATCAATCCCTGCTCCACATCTTGCCCAATGGGAAATGATCTT 1666
    |||||
Db 144084 TCAGTGAATGATATCAATCAATCCCTGCTCCACATCTTGCCCAATGGGAAATGATCTT 144143
OY 1667 TCACCAAGAGCTCACACAGCATTTTCCACAGAGATGC 1703
    |||||
Db 144144 TCACCAAGAGCTCACACAGCATTTTCCACAGAGATGC 144180

RESULT 5
AC138848/c 206943 bp DNA linear HTG 21-JAN-2003
LOCUS Homo sapiens chromosome 5 clone RP11-1301B21, WORKING DRAFT
DEFINITION SEQUENCE, 7 unordered pieces.
ACCESSION AC138848
VERSION AC138848.1 GI:27805260
KEYWORDS HTG: HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 206943)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
PUBLISHED 2 (bases 1 to 206943)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (21-JAN-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 2719263
Center clone name: RPCR-11_1301B21
-----
Summary Statistics
Consensus quality: 202039 bases at least Q40
Consensus quality: 202804 bases at least Q30
Consensus quality: 203337 bases at least Q20
Estimated insert size: 175000; agarose-fp estimation
Estimated insert size: 206343; sum-of-coverage estimation
Quality coverage: 12.96 in Q20 bases; agarose-fp estimation
Quality coverage: 12.96 in Q20 bases; sum-of-coverage estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 7 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1 1219: contig of 1219 bp in length
* 1220 1319: gap of unknown length
* 1320 2493: contig of 1174 bp in length
* 2494 2593: gap of unknown length
* 3665 3766: contig of 1073 bp in length
* 3667 3766: gap of unknown length

```

```

* 3767 5167: contig of 1401 bp in length
* 5168 5267: gap of unknown length
* 5268 12818: contig of 7551 bp in length
* 12819 36169: gap of unknown length
* 36170 206943: contig of 170674 bp in length.
* 36270 Location/Qualifiers
      1. 206943
        /organism="Homo sapiens"
        /mol_type="genomic DNA"
        /db_xref="taxon:9606"
        /chromosome="5"
        /clone="RP11-1301B21"
        /clone_11b="RPC1 human BAC library 11"

```

```

BASE COUNT 56312 a 44122 c 46163 g 59225 t 1121 others
ORIGIN

```

```

Query Match 78.2% Score 1597; DB 2; Length 206943;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

QY 107 AAGGCTCCACCTTACAGACATGCTACTGCTTAAATAAAGAGCTCTGCTGCTG 166
Db 125368 AAGGCTCCACCTTACAGACATGCTACTGCTTAAATAAAGAGCTCTGCTGCTG 125309
QY 167 GTTCCAGATGGGCGATCTGGGCTCTATCTCCACATGTACAGCCACATCAGCTCC 226
Db 125308 GTTCCAGATGGGCGATCTGGGCTCTATCTCCACATGTACAGCCACATCAGCTCC 125249
QY 227 CTGTCTATGAAGGACAGCCGAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 286
Db 125248 CTGTCTATGAAGGACAGCCGAGCGCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 125189
QY 287 GGGCTCTCTTTTGGGGGAGCTTTTGGGAGAGCAGCCAAATGTTTCTACCTGATGAG 346
Db 125188 GGGCTCTCTTTTGGGGGAGCTTTTGGGAGAGCAGCCAAATGTTTCTACCTGATGAG 125129
QY 347 CCGGCGTGGAGCGTGTGATGACCTTCAAGCAGAGCAGCCGCTGGATGTGACATGGCT 406
Db 125128 CCGGCGTGGAGCGTGTGATGACCTTCAAGCAGAGCAGCCGCTGGATGTGACATGGCT 125069
QY 407 GTGCGGATGTGATACGGGCGCTCTTGTGGACATGAGCGCTTGTGATGCTTACATG 466
Db 125068 GTGCGGATGTGATACGGGCGCTCTTGTGGACATGAGCGCTTGTGATGCTTACATG 125009
QY 467 GAACCTGTGCTCCCGGAGACAGTCCCTTTTCAGTGGAGAAACAGCCGGGCGCTGTGT 526
Db 125008 GAACCTGTGCTCCCGGAGACAGTCCAGCTTTTCAGTGGAGAAACAGCCGGGCGCTGTGT 124949
QY 527 TCTGACCTGCTGTGATCATCCCAAGATGAATCAATCCCGGGGCTCACTGAGG 586
Db 124948 TCTGACCTGCTGTGATCATCCCAAGATGAATCAATCCCGGGGCTCACTGAGG 124889
QY 587 CTCCTGTGAGTCAACAGCCCTTTGAGGTGTGAGAGGCGTGGCCTCTCAAGCCAC 646
Db 124888 CTCCTGTGAGTCAACAGCCCTTTGAGGTGTGAGAGGCGTGGCCTCTCAAGCCAC 124829
QY 647 GTGTGTCTAAGAGAGTGGCTTTTCAACCTGACATGCTCTTACCCGCTGTGAAGAC 706
Db 124828 GTGTGTCTAAGAGAGTGGCTTTTCAACCTGACATGCTCTTACCCGCTGTGAAGAC 124769
QY 707 CCGTCCCTCAACCTGATCTGTGACATGCTGTCCGGAGACCCGGGGCGGTTCCTGCTCC 766
Db 124768 CCGTCCCTCAACCTGATCTGTGACATGCTGTCCGGAGACCCGGGGCGGTTCCTGCTCC 124709
QY 767 CGAAGACGACAAAGGAGATCTCATGATTGACAGTGTGATGATGAGGAGCATGAG 826
Db 124708 CGAAGACGACAAAGGAGATCTCATGATTGACAGTGTGATGATGAGGAGCATGAG 124649
QY 827 CAAAACCTCAAGAGAGGAGACCAACCTCTATGTGATGAGAGGTCATCTGCCAAAGCCAG 886
Db 124648 CAAAACCTCAAGAGAGGAGACCAACCTCTATGTGATGAGAGGTCATCTGCCAAAGCCAG 124589

```

```

QY 887 CTGAGATCTACAAGACCATTCAGTCTTGGCCCAAGGCCCTGACAGAACCTTACTGCTT 946
Db 124588 CTGAGATCTACAAGACCATTCAGTCTTGGCCCAAGGCCCTGACAGAACCTTACTGCTT 124529
QY 947 GTGGCGTATGAGAGCAGTGGCTCGAGCCCTGTGGCCCAAGACTTCCGAATGTGAATTC 1006
Db 124528 GTGGCGTATGAGAGCAGTGGCTCGAGCCCTGTGGCCCAAGACTTCCGAATGTGAATTC 124469
QY 1007 GTGGGATGTGAATCTTGGCCCATCTTACAGCCGGGTGATATCAATCACCAGGAG 1066
Db 124468 GTGGGATGTGAATCTTGGCCCATCTTACAGCCGGGTGATATCAATCACCAGGAG 124409
QY 1067 GGCATGGGTGACACAGCTTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGCT 1126
Db 124408 GGCATGGGTGACACAGCTTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGCT 124349
QY 1127 TGGCGCTGTCTTGGCCCATATGAAGAGTTTCTGACTTCAGAAAGCCTGTGGCGATGC 1186
Db 124348 TGGCGCTGTCTTGGCCCATATGAAGAGTTTCTGACTTCAGAAAGCCTGTGGCGATGC 124289
QY 1187 ATGAATTTGCTGGGCTACCCGACAGTCTGATCTGAACAAAGAGAGAAACCTGTCTG 1246
Db 124288 ATGAATTTGCTGGGCTACCCGACAGTCTGATCTGAACAAAGAGAGAAACCTGTCTG 124229
QY 1247 GATCTTGTCTACTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAGGCTTT 1306
Db 124228 GATCTTGTCTACTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAGGCTTT 124169
QY 1307 GCTGCCACCTGTGTGACCTCAGTCACTTCTCTGATGCTGTGAGCTTGTGCTTACAT 1366
Db 124168 GCTGCCACCTGTGTGACCTCAGTCACTTCTCTGATGCTGTGAGCTTGTGCTTACAT 124109
QY 1367 CTCTGACCTTTACTAATGTCTGTGGGTATCACTGAGTGTGATGTTGTCTCAGACGT 1426
Db 124108 CTCTGACCTTTACTAATGTCTGTGGGTATCACTGAGTGTGATGTTGTCTCAGACGT 124049
QY 1427 GCTCAAGCAAGAGACTTTTGTGTCATGCTTGTGTAGAAACAGACTGGGGAGCTT 1486
Db 124048 GCTCAAGCAAGAGACTTTTGTGTCATGCTTGTGTAGAAACAGACTGGGGAGCTT 123989
QY 1487 ATGTGACGACCATCCACCAGTGAAGAGGATGTCCTTCTTCTTCTTGTGATCTT 1546
Db 123988 ATGTGACGACCATCCACCAGTGAAGAGGATGTCCTTCTTCTTCTTGTGATCTT 123929
QY 1547 CCTGTCTGGGAGACTTCAAGAGACTTGTGGCCTGGAGGCTATTAAAGCAGACAGATA 1606
Db 123928 CCTGTCTGGGAGACTTCAAGAGACTTGTGGCCTGGAGGCTATTAAAGCAGACAGATA 123869
QY 1607 TCAGTGAATTTGATCCATTAACCTCCCTGTCCACATCTTGCCCAATGGGAATGATCTT 1666
Db 123868 TCAGTGAATTTGATCCATTAACCTCCCTGTCCACATCTTGCCCAATGGGAATGATCTT 123809
QY 1667 TCACCAAGAGCTCACAGCATTTTTCACAGAGATGC 1703
Db 123808 TCACCAAGAGCTCACAGCATTTTTCACAGAGATGC 123772

```

```

RESULT 6
BD127258
LOCUS 1979 bp DNA linear PAT 18-SEP-2002
DEFINITION Primer for synthesizing full-length cDNA and use thereof.
ACCESSION BD127258
VERSION BD127258.1 GI:23222203
KEYWORDS UP 2002017375-A/2689.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
Koga,H.
REFERENCE
AUTHORS

```

TITLE Primer for synthesizing full-length cDNA and use thereof
 JOURNAL Patent: JP 2002017375-A 2689 22-Jan-2002;
 COMMENT HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/2689
 PD 22-Jan-2002
 PF 07-JUL-2000 JP 2002053172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHIMICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOA
 PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10,
 PC C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT CDS (69)..(1226).
 Location/Qualifiers
 source 1..1979
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 451 a 542 c 474 g 512 t
 ORIGIN
 Query Match 75.7%; Score 1547; DB 6; Length 1979;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

Db 647 CCCCTCCCTCAACCTGCATATCGTGCACCTGTCCGGAGCCCCGGCCGCTTCCGTTG 706
 QY 766 CCGAAGACGACAAAGAGAGATCTCATGATTAAGATCCATGTTGATGGGACAGATGA 825
 Db 707 CCGAAGACGACAAAGAGAGATCTCATGATTAAGATCCATGTTGATGGGACAGATGA 766
 QY 826 GCAAAAACGCAAGAGAGAGACCAACCTACTATGATGATGACAGTCAATCTGCCAAGCA 885
 Db 767 GCAAAAACGCAAGAGAGAGACCAACCTACTATGATGATGACAGTCAATCTGCCAAGCA 826
 QY 886 GCTGAGATCTCAAGACATCCAGTCTTGGCCAAAGGCCCTGCAGAGAGGCTACTGCT 945
 Db 827 GCTGAGATCTCAAGACATCCAGTCTTGGCCAAAGGCCCTGCAGAGAGGCTACTGCT 886
 QY 946 TGTGCGCTATGAGAGACCTGCTGAGCCCTGTGGCCCGAGCTCCGATGTATGAAT 1005
 Db 887 TGTGCGCTATGAGAGACCTGCTGAGCCCTGTGGCCCGAGCTCCGATGTATGAAT 946
 QY 1006 CGTGGGATTGGAATTTTCCCATCTTACAGACTGGTGATTAACATCAACCCGAGGCAA 1065
 Db 947 CGTGGGATTGGAATTTTCCCATCTTACAGACTGGTGATTAACATCAACCCGAGGCAA 1006
 QY 1066 GGGCATGGGTGACCAAGCTTTCCACAAATGCGAGAGATGCCCTTAATGTCCAGGC 1125
 Db 1007 GGGCATGGGTGACCAAGCTTTCCACAAATGCGAGAGATGCCCTTAATGTCCAGGC 1066
 QY 1126 TTGGCGCTGATCTTGGCCCTATGAAAAGTTTCTGACATTCAGAAAACCTGTGGCATGC 1185
 Db 1067 TTGGCGCTGATCTTGGCCCTATGAAAAGTTTCTGACATTCAGAAAACCTGTGGCATGC 1126
 QY 1186 CATGAATTTGTTGGGTACCCGACGTCAAGTCTGAACAAGAAAGAAACCTGTGGCT 1245
 Db 1127 CATGAATTTGTTGGGTACCCGACGTCAAGTCTGAACAAGAAAGAAACCTGTGGCT 1186
 QY 1246 GGAATTTGTTGTTACCTGAGTCTGCTCAGCAAAATCCACTAAGAGGTTGAGAAAGCTT 1305
 Db 1187 GGAATTTGTTGTTACCTGAGTCTGCTCAGCAAAATCCACTAAGAGGTTGAGAAAGCTT 1246
 QY 1306 TGTGCGCACTGGGTGTCAGCCCTGACAGTCACTTCTGTAATGCTTCTGAGCCTTGCCTACA 1365
 Db 1247 TGTGCGCACTGGGTGTCAGCCCTGACAGTCACTTCTGTAATGCTTCTGAGCCTTGCCTACA 1306
 QY 1366 TGTGCGCACTGGGTGTCAGCCCTGACAGTCACTTCTGTAATGCTTCTGAGCCTTGCCTACA 1425
 Db 1307 TGTGCGCACTGGGTGTCAGCCCTGACAGTCACTTCTGTAATGCTTCTGAGCCTTGCCTACA 1366
 QY 1426 TGTGCGCACTGGGTGTCAGCCCTGACAGTCACTTCTGTAATGCTTCTGAGCCTTGCCTACA 1485
 Db 1367 TGTGCGCACTGGGTGTCAGCCCTGACAGTCACTTCTGTAATGCTTCTGAGCCTTGCCTACA 1426
 QY 1486 TATGAGAGACGACATCCACCAAGTGAAGAGGATATGCTCTCTCTTCTTCTGATCT 1545
 Db 1427 TATGAGAGAGACGACATCCACCAAGTGAAGAGGATATGCTCTCTCTTCTTCTGATCT 1486
 QY 1546 TCTGTCTGAGGAGACTTCAAGAGACTTGTGGCTGGAGAGGCTATTAACAGAGACAGT 1605
 Db 1487 TCTGTCTGAGGAGACTTCAAGAGACTTGTGGCTGGAGAGGCTATTAACAGAGACAGT 1546
 QY 1606 ATCAGTGAATGATCATAACTCCCTGTCACATCTTGGCCATGGGAGATGATCT 1665
 Db 1547 ATCAGTGAATGATCATAACTCCCTGTCACATCTTGGCCATGGGAGATGATCT 1606
 QY 1666 TTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
 Db 1607 TTCACCAAGAGCTCACAGCATTTTCCACAGAGATGC 1644
 RESULT 7
 AK074746 1979 bp mRNA linear PRI 03-SEP-2002
 LOCUS AK074746
 DEFINITION Homo sapiens cDNA FLJ90265 fis, clone NTBM4001325, weakly similar
 to CHONDROITIN 6-SULFOTRANSFERASE (EC 2.8.2.17).
 ACCESSION AK074746

```

VERSION      AK074746.1 GI:22760388
KEYWORDS     oligo capping; fls (full insert sequence).
SOURCE       Homo sapiens (human)
ORGANISM     Eukaryotes; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE    1
AUTHORS      Isogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T.,
              Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
              Kawai-Hito,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
              Kojima,S., Nagahara,K., Masuno,Y., Ono,T., Okano,K., Yoshikawa,Y.,
              Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
              Ninomiya,K.
TITLE        NEDO human cDNA sequencing project
JOURNAL      Unpublished
REFERENCE    2 (bases 1 to 1979)
AUTHORS      Isogai,T. and Otsuki,T.
TITLE        Direct Submission
JOURNAL      Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
              Genomics Laboratory; 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
              (E-mail:genomices@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
              NEDO human cDNA sequencing project supported by Ministry of
              Economy, Trade and Industry of Japan; cDNA full insert sequencing:
              Research Association for Biotechnology; cDNA library construction:
              Institute of Medical Science, University of Tokyo, Laboratory of
              Genome Structure and Human Genome Center; cDNA 5'- & 3'-end one pass
              sequencing and clone selection: Helix Research Institute (supported
              by Japan Key Technology Center etc.).

FEATURES
  source
    1..1979
    /organism="Homo sapiens"
    /mol_type="mRNA"
    /db_xref="taxon:9606"
    /clone="NT2RM4.001325"
    /cell_line="NT2"
    /cell_type="teratocarcinoma"
    /clone_lib="NT2RM4"
    /note="cloning vector: pME18SFL3-mRNA from uninduced NT2
    neuronal precursor cells"
    69..1229
    /note="unnamed protein product"
    /codon_start=1
    /protein_id="BAC11177.1"
    /db_xref="GI:22760389"
    /translation="MLPKKKLLFLVSOMATILAFEHYSHNISLSKAPDRBH
    VILSSMRSSSFVGLFGHDPYFVLEMEBAHVMWFKSTAMLMHAYRDLIRAF
    LCDMSYFDATMEBPGRQSSLPWENSRALCSAPACITIFODELIPRRLCSQDP
    FEVEERACRSYSHVLEKEVEFNQSLYPLKDPSLMHLVHVRPRAVERSERK
    GDLMDNRIYMGHQLKKEDQPYVMQVTCOSLEIYTIQSLPRALOERYLIVY
    EDLARAIVAQTSRMFEVGLFPLHLOTVHNITRGGMGDHAFHTVARADALNSQAM
    RWSLPYKVRRLQKAGDAMNLGYSRHRVRESDQRNLLDLLSTWYVPDQIH"

CDS
  451 a      542 c      474 g      512 t
  RIGIN
  Query Match      75.7%; Score 1547; DB 9; Length 1979;
  Best Local Similarity 99.9%; Pred. No. 0;
  Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

  QY 106 CAAGGCTTCCACTTTCAGCACAATGCTACGCGCCCAAAAAATGAAAGCTCTGCTGTTTCT 165
  DB 47 CAAGGCTTCCACTTTCAGCACAATGCTACGCGCCCAAAAAATGAAAGCTCTGCTGTTTCT 106
  QY 166 GGTTCACAGATGGCCATCTTGCTCTATCTTCACATGATGACGACCAACATCAGCTC 225
  DB 107 GGTTCACAGATGGCCATCTTGCTCTATCTTCACATGATGACGACCAACATCAGCTC 166
  QY 226 CCGTCTATGAAGGACAGCCGACGACATGACAGCTGCTGCTGCTCTGCTGCGCTC 285
  DB 167 CCGTCTATGAAGGACAGCCGACGACATGACAGCTGCTGCTGCTGCTGCGCTC 226
  QY 286 TGCTCTCTCTTTTGTGGGAGAGCTTTTGGGACAGACACCGAATGTTTCTTACCTGATGGA 345
  DB 227 TGCTCTCTCTTTTGTGGGAGAGCTTTTGGGACAGACACCGAATGTTTCTTACCTGATGGA 286

```

```

  QY 346 GCCCGCTGGCAGCTGTGATGACCTTCCAAAGCAGACGCGCTGATGCTGCAATGGC 405
  DB 287 GCCCGCTGGCAGCTGTGATGACCTTCCAAAGCAGACGCGCTGATGCTGCAATGGC 346
  QY 406 TGTGCGGATGATGATAGGCGCGCTCTTGTGACATGAGCTCTTTGATGGCTACAT 465
  DB 347 TGTGCGGATGATGATAGGCGCGCTCTTGTGACATGAGCTCTTTGATGGCTACAT 406
  QY 466 GGAACCTGTGTCCTCCGAGACAGCTCAGCTCTTTCAGTGGAGAAACAGCGGCGCTGTG 525
  DB 407 GGAACCTGTGTCCTCCGAGACAGCTCAGCTCTTTCAGTGGAGAAACAGCGGCGCTGTG 466
  QY 526 TTTCGACACCTGCGTGCATCATCATCCACAAAGATGAATATATCCCGCGGCTCACTGAG 585
  DB 467 TTTCGACACCTGCGTGCATCATCATCCACAAAGATGAATATATCCCGCGGCTCACTGAG 526
  QY 586 GCTCCTGTGACATGACAGCCCTTGAAGTGTGAGAGAGAGCTGCGCTCTACAGCA 645
  DB 527 GCTCCTGTGACATGACAGCCCTTGAAGTGTGAGAGAGAGCTGCGCTCTACAGCA 586
  QY 646 CGTGCTGCTCAAGAGAGTGGCTCTTCAACCTGCACTCCCTTACCCGCTGTGAAAGA 705
  DB 587 CGTGCTGCTCAAGAGAGTGGCTCTTCAACCTGCACTCCCTTACCCGCTGTGAAAGA 646
  QY 706 CCCCTCCCTCAACCTGCATATCGACACCTGCTGCGGAGACCCCGGCGCTTCCGCTTC 765
  DB 647 CCCCTCCCTCAACCTGCATATCGACACCTGCTGCGGAGACCCCGGCGCTTCCGCTTC 706
  QY 766 CCGAAGACGCAAAAGGAGATCTCATGATGACAGTGCATGATGAGGGGAGCATGA 825
  DB 707 CCGAAGACGCAAAAGGAGATCTCATGATGACAGTGCATGATGAGGGGAGCATGA 766
  QY 826 GCAAAACTCAAGAGAGAGACCAACCTACTATGTGATGAGTCACTTCCAAAGCA 885
  DB 767 GCAAAACTCAAGAGAGAGACCAACCTACTATGTGATGAGTCACTTCCAAAGCA 826
  QY 886 GCTGAGATGTACAAAGACATCCAGTCCCTTCCCAAGGCGCTGAGAGAGCTACCTGCT 945
  DB 827 GCTGAGATGTACAAAGACATCCAGTCCCTTCCCAAGGCGCTGAGAGAGCTACCTGCT 886
  QY 946 TGTGCGCTATGAGAGACCTGCTGAGCCCTGTGAGCCAGACTTCCGAATGATGAAT 1005
  DB 887 TGTGCGCTATGAGAGACCTGCTGAGCCCTGTGAGCCAGACTTCCGAATGATGAAT 946
  QY 1006 CGTGGAATTTGAATTTTGGCCCATCTTTCAGACCTGGGTGATTAACATCACCCGAGCAA 1065
  DB 947 CGTGGAATTTGAATTTTGGCCCATCTTTCAGACCTGGGTGATTAACATCACCCGAGCAA 1006
  QY 1066 GGGCATGGGAGACAGCTTTCACACAAATGGCAGGAGATGCCCTTAATGTCGCCAGGC 1125
  DB 1007 GGGCATGGGAGACAGCTTTCACACAAATGGCAGGAGATGCCCTTAATGTCGCCAGGC 1066
  QY 1126 TTGGCGCTGTGCTTTGGCCATATGAAAAAGTTTCTGACTTCAGAAAGCTGTGCGATGC 1185
  DB 1067 TTGGCGCTGTGCTTTGGCCATATGAAAAAGTTTCTGACTTCAGAAAGCTGTGCGATGC 1126
  QY 1186 CATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAAACAGAAAGAAACCTGTGCT 1245
  DB 1127 CATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAAACAGAAAGAAACCTGTGCT 1186
  QY 1246 GGATCTCTGCTCTACCTGAGACTGCTGCTGAGCAATCAGCTTAAGAGGTTGAGAAAGCTT 1305
  DB 1187 GGATCTCTGCTCTACCTGAGACTGCTGCTGAGCAATCAGCTTAAGAGGTTGAGAAAGCTT 1246
  QY 1306 TGCTGACACCTGATGTAGGCTCAGTCACTTCTCTGAATGCTTCTAGGCTTGCCTACA 1365
  DB 1247 TGCTGACACCTGATGTAGGCTCAGTCACTTCTCTGAATGCTTCTAGGCTTGCCTACA 1306
  QY 1366 TCTCTGAGCTTAACTACATGCTGTGGGATATCATCATGATGATGATGATGATGATGATGAT 1425
  DB 1307 TCTCTGAGCTTAACTACATGCTGTGGGATATCATCATGATGATGATGATGATGATGATGAT 1366

```


Query Match	Best Local Similarity	75-7%: 99.9%	Score 1547	DB 6	Length 2032
Matches 1597; Conservative	99.9%	0; Mismatches 1; Indels 0; Gaps 0;			
106 CAAGGTCCTTCACCTTACAGACAATGACTCCTAAATAATGAAGCTCCTGCTGTTCT	165				
160 CAAGGTCCTTCACCTTACAGACAATGACTCCTAAATAATGAAGCTCCTGCTGTTCT	219				
166 GGTTCCTCCAGATGAGCCATCTTGGCTCTATTCTTCCATGTAGACCAACAATCAGCTC	225				
220 GGTTCCTCCAGATGAGCCATCTTGGCTCTATTCTTCCATGTAGACCAACAATCAGCTC	279				
226 CCTGTCTATGAGGACACAGCCCGGAGCGCATGACAGTCTGCTCTGTCTTCTTGGCGCTC	285				
280 CCTGTCTATGAGGACACAGCCCGGAGCGCATGACAGTCTGCTCTGTCTTCTTGGCGCTC	339				
286 TGCGTCTTCTTTTGTGGGGCAGCTTTTGTGGGACAGCCACCAATGTTTCTTCACTGATGGA	345				
340 TGCGTCTTCTTTTGTGGGGCAGCTTTTGTGGGACAGCCACCAATGTTTCTTCACTGATGGA	399				
346 GCCCGCCTGGCAGCTGTGATGACCTTCAAGCAGACAGCAGCCGCTGATGCTGACATGGC	405				
400 GCCCGCCTGGCAGCTGTGATGACCTTCAAGCAGACAGCAGCCGCTGATGCTGACATGGC	459				
406 TGTCGGGAGATGATAGGGGCGCTCTTGTGGGACATGAGCGCTTGTGATGCTACAT	465				
460 TGTCGGGAGATGATAGGGGCGCTCTTGTGGGACATGAGCGCTTGTGATGCTACAT	519				
466 GGAACCTGTGTCCTCCGGAGACAGTCCAGCTCTTTCAGTGGGAGAAACAGCCGGGCTCTGTG	525				
520 GGAACCTGTGTCCTCCGGAGAGAGTCCAGCTCTTTCAGTGGGAGAAACAGCCGGGCTCTGTG	579				

OY	526	TTCTGACACCTGCGTGCATCATCCACCAAGATGAATCATCTCCCGGGCTCACTGAG	585
Db	580	TTCTGACACTGCTGTGACATCATCCCAAGATGAATCATCTCCCGGGCTCACTGAG	639
OY	586	GCTCCCTGTGCAGTCAACAGCCCTTTGAGGTGGTGGAGAAAGGCTGCGCTCTCAAGCA	645
Db	640	GCTCCCTGTGCAGTCAACAGCCCTTTGAGGTGGTGGAGAAAGGCTGCGCTCTCAAGCA	699
OY	646	CGTGGTGTCAAGGAGGTGGCCTTCTTCAACCTGCAGTCCCTCTACCGCTGCTGAAGA	705
Db	700	CGTGGTGTCAAGGAGGTGGCCTTCTTCAACCTGCAGTCCCTCTACCGCTGCTGAAGA	759
OY	706	CCCCCTCCCTCAACCTGCATTCGTGACACCTGGTCCGGGACCCCGGGGCCGTTC	765
Db	760	CCCCCTCCCTCAACCTGCATTCGTGACACCTGGTCCGGGACCCCGGGGCCGTTC	819
OY	766	CCGAGAACGCACAAAGGAGATCTCATGATTTGACAGTTCGATTTGTATGGGGCAGCATGA	825
Db	820	CCGAGAACGCACAAAGGAGATCTCATGATTTGACAGTTCGATTTGTATGGGGCAGCATGA	879
OY	826	GCAAAAACTCAAAGAAAGGAGACCAACCTACTATGTATGTGAGAGTCAATCTCCCAAAGCA	885
Db	880	GCAGAAACTCAAAGAAAGGAGACCAACCTACTATGTATGTGAGAGTCAATCTCCCAAAGCA	939
OY	886	GCTGGAGATTCACAAAGACATTCAGTCCCTTCCCAAAGGCCCTGGAGGAAACCTCACTGCT	945
Db	940	GCTGGAGATTCACAAAGACATTCAGTCCCTTCCCAAAGGCCCTGGAGGAAACCTCACTGCT	999
OY	946	TGTGCGCTATGAGGACCTGCGCTCGAGCCCTGTGGCCCGCAGCTTCCGGAATGTATGAATT	1005
Db	1000	TGTGCGCTATGAGGACCTGCGCTCGAGCCCTGTGGCCCGCAGCTTCCGGAATGTATGAATT	1059
OY	1006	CGTGGGATTTGGAATTTCTTGCCCAATCTTCAACACTGGGTGCATTAACATCACCCAGGCAA	1065
Db	1060	CGTGGGATTTGGAATTTCTTGCCCAATCTTCAACACTGGGTGCATTAACATCACCCAGGCAA	1119
OY	1066	GGGCGATGGGCTACCAACGCTTTCACACACAAATGGCCAGGGATGCCCTTAATGTCTCCAGGC	1125
Db	1120	GGGCGATGGGCTACCAACGCTTTCACACACAAATGGCCAGGGATGCCCTTAATGTCTCCAGGC	1179
OY	1126	TTGGCGCGTGTCTTGGCCCTATGAAAAGGTTTCTCGACTTCAGAAAGGCTTGCGCGATGC	1185
Db	1180	TTGGCGCGTGTCTTGGCCCTATGAAAAGGTTTCTCGACTTCAGAAAGGCTTGCGCGATGC	1239
OY	1186	CATGAATTTGCTGGGCTACCGCCACAGTCAGATCTGAACAGAACAGAAACCTGTGCT	1245
Db	1240	CATGAATTTGCTGGGCTACCGCCACAGTCAGATCTGAACAGAACAGAAACCTGTGCT	1299
OY	1246	GGATTCCTGTCTACCTGGACGTGTCGCCGACCAATCCACTAAGAGGGTTAGAAAGGCTT	1305
Db	1300	GGATTCCTGTCTACCTGGACGTGTCGCCGACCAATCCACTAAGAGGGTTAGAAAGGCTT	1359
OY	1306	TGCTGCGCACCTGCGTGCAGCCTCAAGTCACCTTCTCTGATGCTTCTGAGCCTTGCCTACA	1365
Db	1360	TGCTGCGCACCTGCGTGCAGCCTCAAGTCACCTTCTCTGATGCTTCTGAGCCTTGCCTACA	1419
OY	1366	TCTCTGAGCCTTAACTACATGTCTGTGGGTATCAACACTGAGTGTGAGTTGTCTCCACAG	1425
Db	1420	TCTCTGAGCCTTAACTACATGTCTGTGGGTATCAACACTGAGTGTGAGTTGTCTCCACAG	1479
OY	1426	TGCTCAGACGAAGAGACTTTGTGTGCCATGCTTTGTGCTTGAANAAGACGTGGGGAACCT	1485
Db	1480	TGCTCAGACGAAGAGACTTTGTGTGCCATGCTTTGTGCTTGAANAAGACGTGGGGAACCT	1539
OY	1486	TATGTGAGACAGCATCCACACCACTGAAACAGGGTATTTGCTCTTCTTCTTCTTGATCT	1545
Db	1540	TATGTGAGACAGCATCCACACCACTGAAACAGGGTATTTGCTCTTCTTCTTCTTGATCT	1599
OY	1546	TCTCTGTCTGGGAGACTTCAGAGACTTTGTGGCTGGAGGCTTATTAAGCAGCAGACAGT	1605
Db	1600	TCTCTGTCTGGGAGACTTCAGAGACTTTGTGGCTGGAGGCTTATTAAGCAGCAGACAGT	1659

QY	1606	ATCAGTGAATTGAAATTCATTAACGCCCTGTCACATCTGTCCCAATGGGAATGAGTCT	1606
Db	1660	ATCAGTGAATTAATGATTCATTAACCTCCCTGTCACATCTGTCCCAATGGGAATGATCT	1719
QY	1666	TTTCAACAAGAGCTCACCAGCATTTTCCACAGAAATGC	1703
Db	1720	TTTCAACAAGAGCTCACCAGCATTTTCCACAGAAATGC	1757
RESULT 9			
LOCUS	BD134772	2032 bp	DNA
DEFINITION	BD134772		Linear
ACCESSION	BD134772.1		PAT 18-SEP-2002
VERSION	BD134772.1		
KEYWORDS	JP 2002507409-A/1.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.		
AUTHORS	Bistup,A., Rosen,S.D., Tangemann,K. and Hemmerlich,S.		
TITLE	GLYCOSYLsulfoltransferase-3		
JOURNAL	Patent: JP 2002507409-A 12-MAR-2002;		
COMMENT	THE REGENES OF THE UNIVERSITY OF CALIFORNIA, SYNTAX INC		
	OS Homo sapiens (human)		
	PN JP 2002507409-A/1		
	PD 12-MAR-2002		
	PF 26-FEB-1999 JP 2000537979		
	PR 20-MAR-1998 US 09/045284,12-NOV-1998 US 09/190911 PI		
	ANNETTE BISTUP, STEVEN D ROSEN, KIRSTEN TANEMANN, STEFAN PI		
	HEMMERLICH		
	PC C12N9/10,A01K67/027,A61K45/00,A61K48/00,A61P29/00,A61P37/06,		
	PC A61P43/00,		
	PC C12N1/15,C12N1/21,C12N5/10,C12N15/09,C1201/48,C12N5/00,C12N15/15,		
	PC 00		
	CC GLYCOSYLsulfoltransferase-3		
	FH Key location/Qualifiers		
	FT source 1..2032		
FEATURES			
source	location/Qualifiers		
	1..2032		
	/organism="Homo sapiens"		
	/mol_type="genomic DNA"		
	/db_xref="taxon:9606"		
BASE COUNT	468 a 569 c 490 g 505 t		
ORIGIN			
	</		

Dd	460	TTGTGGGGATGTGATACGGGCGGCTTCTTGTGTGGACATGGCGCTTTGTGATGCTCAT	519
Qy	466	GGAACTGTGTCCTCCGGAGACAGTCCAGGCTTTTCAGTGGGAGAAACCGGGGCTCTGTG	525
Dd	520	GGAACTGTGTCCTCCGGAGACAGTCCAGGCTTTTCAGTGGGAGAAACCGGGGCTCTGTG	579
Qy	526	TTTGTGACCTGCTGTGTGATCATTCATCCACAAAGATGAATCATCCCGGGGCTACTGTGAG	585
Dd	580	TTTGTGACCTGCTGTGTGATCATTCATCCACAAAGATGAATCATCCCGGGGCTACTGTGAG	639
Qy	586	GCTCCTGTGACGTCAACAGCCCTTGTGAGTGTGTGAGAAAGGCTCGGCTCTACAGCA	645
Dd	640	GCTCCTGTGACGTCAACAGCCCTTGTGAGTGTGTGAGAAAGGCTCGGCTCTACAGCA	699
Qy	646	CGTGTGTGTCGAAGGAGGTGCGCTTCTTCACACTGACGCTCTACCCGCTGCTGAAGA	705
Dd	700	CGTGTGTGTCGAAGGAGGTGCGCTTCTTCACACTGACGCTCTACCCGCTGCTGAAGA	759
Qy	706	CCCTCCCTCAACTGCATATGTGTGACACTGTGTCCGGGACCCCGGGGCTGTTCGTTTC	765
Dd	760	CCCTCCCTCAACTGCATATGTGTGACACTGTGTCCGGGACCCCGGGGCTGTTCGTTTC	819
Qy	766	CCGGAACGCGCAAAAGGAGATCTCATGATTGTGACAGTGTGCACTGTGTGATGGGGGACGATGA	825
Dd	820	CCGGAACGCGCAAAAGGAGATCTCATGATTGTGACAGTGTGCACTGTGTGATGGGGGACGATGA	879
Qy	826	GCAAAAATCTCAAGAGAGAGACCAACCTACTATGTGTATGATGAGTCTTGTCCAAAGCA	885
Dd	880	GCAGAAATCTCAAGAGAGAGACCAACCTACTATGTGTATGATGAGTCTTGTCCAAAGCA	939
Qy	886	GCTGGAGATCTACAAAGACATCCAGTCTTGTCCCAAGGCTCTGCAAGAGAGCTTACCTGTCT	945
Dd	940	GCTGGAGATCTACAAAGACATCCAGTCTTGTCCCAAGGCTCTGCAAGAGAGCTTACCTGTCT	999
Qy	946	TGTGGGCTATAGAGACCTGTGCTGAGGCCCTGTGGGCCAGACTTCCGAAATGTATGATT	1005
Dd	1000	TGTGGGCTATAGAGACCTGTGCTGAGGCCCTGTGGGCCAGACTTCCGAAATGTATGATT	1059
Qy	1006	CGTGGGATTTGAATTCCTTGCCCAATCTTTCAGACCTGTGGTGTGCTAAATCACCCAGGCAA	1065
Dd	1060	CGTGGGATTTGAATTCCTTGCCCAATCTTTCAGACCTGTGGTGTGCTAAATCACCCAGGCAA	1119
Qy	1066	GGGCGATGGGTGACACAGCTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGC	1125
Dd	1120	GGGCGATGGGTGACACAGCTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGC	1179
Qy	1126	TTGGGGCTGTGCTTGGCCCTATGAAAAGGCTTTCGACTGTAGAAAAGCTGTGGCGATGC	1185
Dd	1180	TTGGGGCTGTGCTTGGCCCTATGAAAAGGCTTTCGACTGTAGAAAAGCTGTGGCGATGC	1239
Qy	1186	CATGAATTTGCTGGGCTACCGCCACAGTGTGATCTGAAACAAAGAGAAACCTGTGTCT	1245
Dd	1240	CATGAATTTGCTGGGCTACCGCCACAGTGTGATCTGAAACAAAGAGAAACCTGTGTCT	1299
Qy	1246	GGATCTTGTGTCTACCTGTGACTGTTCCTGACAAATCCACTAAGAGGTTGAGAAGCTT	1305
Dd	1300	GGATCTTGTGTCTACCTGTGACTGTTCCTGACAAATCCACTAAGAGGTTGAGAAGCTT	1359
Qy	1306	TGTGCGCACCGGTGTGACCTCGCTCAAGTCACTTTCCTGAATGCTTGTAGGCTTGTGCCTACA	1365
Dd	1360	TGTGCGCACCGGTGTGACCTCGCTCAAGTCACTTTCCTGAATGCTTGTAGGCTTGTGCCTACA	1419
Qy	1366	TCTGTGAGACCTTAACTACATGTCTGTGGGTATACACACTGAGTGTGAGTGTGTGCTACACG	1425
Dd	1420	TCTGTGAGACCTTAACTACATGTCTGTGGGTATACACACTGAGTGTGAGTGTGTGCTACACG	1479
Qy	1426	TGCTTCAGACAGAGGACTTTTGTGTCCATGTCTGTCTAGAAAAACAGACTGGGAACT	1485
Dd	1480	TGCTTCAGACAGAGGACTTTTGTGTCCATGTCTGTCTAGAAAAACAGACTGGGAACT	1539
Qy	1486	TATGTGAGACACATCCACAGTGAAGAGGATTTGCTTCTTCTTCTTGTGATCT	1545

Db	1540	TATGTGACGACACATCCACACAGTGAAGAAGGGGATATGCTTCTCTCTTTTCCTGATCT	1599
0y	1546	TCCTGTCTGGGACACACTTCACAGACTTTGTGTGCGTCGAGGCCCTATTAAAGCAGACACAGT	1605
Db	1600	TCCTGTCTGGGACACACTTCACAGACTTTGTGTGCGTCGAGGCCCTATTAAAGCAGACACAGT	1659
0y	1606	ATCATGTGAATTGATCCATTAACCTCCCTGTCACATCTTGCCTCCCAATGGGGAAATGATCT	1665
Db	1660	ATCATGTGAATTGATCCATTAACCTCCCTGTCACATCTTGCCTCCCAATGGGGAAATGATCT	1719
Db	1660	ATCATGTGAATTGATCCATTAACCTCCCTGTCACATCTTGCCTCCCAATGGGGAAATGATCT	1719
0y	1666	TTTCACCAAAAGAGCTCACCAGCATTTTTCACAGATGC	1703
Db	1720	TTTCACCAAAAGAGCTCACCAGCATTTTTCACAGATGC	1757
RESULT 10			
AF131235		2032 bp	mRNA
LOCUS			linear
DEFINITION			PRI 10-AUG-1999
ACCESSION	AF131235		
VERSION	AF131235		
KEYWORDS	AF131235.1	GI:4927113	
SOURCE			
ORGANISM			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			
REFERENCE			
AUTHORS			
TITLE			
JOURNAL			
MEDLINE			
PUBMED			

Db	160	CAAGGCTCTTCCACTTCACACACATGCTCTACTGCTCTAAAAAAATGAAGCTCTGCTGTTTCT	213
OY	166	GGTTTCCAGATGGCCATCTTGCGCTATTTCTTCCATGTACAGCCACACATCATGCTC	225
Db	220	GGTTTCCAGATGGCCATCTTGCGCTATTTCTTCCATGTACAGCCACACATCATGCTC	279
OY	226	CCTGCTATGAAAGGACACAGCCCGACAGCATGACGTGCTGCTTCTTCCCTGGCGCTC	285
Db	280	CCTGCTATGAAAGGACACAGCCCGACAGCATGACGTGCTGCTTCTTCCCTGGCGCTC	339
OY	286	TGGCTCTTCTTTTGTGGGGGAGACTTTTGTGGGAGACCCAGATGTTTTCTACTCTATGGA	345
Db	340	TGGCTCTTCTTTTGTGGGGGAGACTTTTGTGGGAGACCCAGATGTTTTCTACTCTATGGA	399
OY	346	GCCCGCCTGGACACGTGTGATGACCTTCAAGACAGACACCGCTGGATGCTGCACATGGC	405
Db	400	GCCCGCCTGGACACGTGTGATGACCTTCAAGACAGACACCGCTGGATGCTGCACATGGC	459
OY	406	TGTGCGGGATCTGATACGGGCGCTTCTTGTGGCACAATGACGCTTTTGATGCTTACAT	465
Db	460	TGTGCGGGATCTGATACGGGCGCTTCTTGTGGCACAATGACGCTTTTGATGCTTACAT	519
OY	466	GGAACCTGGTCCCCGGAGACAGTCCACGCTCTTTAGTGGGAGACAGCGCGGCGCTGNG	525
Db	520	GGAACCTGGTCCCCGGAGACAGTCCACGCTCTTTAGTGGGAGACAGCGCGGCGCTGNG	579
OY	526	TTTGTGCACCTGCTGTGACATCATCCACAAGATGAATTCATCCCGGGGCTCACTGGAC	585
Db	580	TTTGTGCACCTGCTGTGACATCATCCACAAGATGAATTCATCCCGGGGCTCACTGGAC	639
OY	586	GCTCCTGTGCAGTCAACAGCCCTTTGAGGTGTGGAGAAAGGCTTCCGCTCTTACAGCCA	645
Db	640	GCTCCTGTGCAGTCAACAGCCCTTTGAGGTGTGGAGAAAGGCTTCCGCTCTTACAGCCA	699
OY	646	CGTGGGTCTCAAGGGAGTGGCGCTTCTTCAACCTGAGTCCGCTCACCGGCTGTCAAAAG	705
Db	700	CGTGGGTCTCAAGGGAGTGGCGCTTCTTCAACCTGAGTCCGCTCACCGGCTGTCAAAAG	759
OY	706	CCCTCTCCCTCAACCTGTATCTGTGACACTGGTCCGGGAGCCCGCGGCGTGTCCGTTTC	765
Db	760	CCCTCTCCCTCAACCTGTATCTGTGACACTGGTCCGGGAGCCCGCGGCGTGTCCGTTTC	819
OY	766	CCGAGAACGCACAAAGGAGATCTCTATGATGTGACAGTGGCATTTGATGTGGGCGACATGA	825
Db	820	CCGAGAACGCACAAAGGAGATCTCTATGATGTGACAGTGGCATTTGATGTGGGCGACATGA	879
OY	826	GCAAAAATCTCAAGAAAGAGAGCAACCTTACTATGTGATGAGCATCTGTGCAAAAGCA	885
Db	880	GCAAAAATCTCAAGAAAGAGAGCAACCTTACTATGTGATGAGCATCTGTGCAAAAGCA	939
OY	886	GCTGAGATCTACAAGACCATTCAGTCTTGTGCCCAAGGCGCTGTGAGAAAGCTACTGCTCT	945
Db	940	GCTGAGATCTACAAGACCATTCAGTCTTGTGCCCAAGGCGCTGTGAGAAAGCTACTGCTCT	999
OY	946	TGTGCGCTATAGAGACCTGTGCTCGAGGCCCTGTGGCCCAAGACTTCCCGAATGTATGAAAT	1007
Db	1000	TGTGCGCTATAGAGACCTGTGCTCGAGGCCCTGTGGCCCAAGACTTCCCGAATGTATGAAAT	1055
OY	1006	CGTGGGATTTGAAATTTCTTGCCCAATCTTCAGACCTGGGTGCATATCATCACCCGAGCAA	1065
Db	1060	CGTGGGATTTGAAATTTCTTGCCCAATCTTCAGACCTGGGTGCATATCATCACCCGAGCAA	1115
OY	1066	GGGCGATGGGTGACACGCTTTTCCACACAAATGCGAGGGATGCGCTTAATGTCTCCACAGC	1122
Db	1120	GGGCGATGGGTGACACGCTTTTCCACACAAATGCGAGGGATGCGCTTAATGTCTCCACAGC	1177
OY	1126	TTTGGCGCTGTGCTTTTGCCCTATGAAAAGTTTCTGACTTCAGAAAAGCTGTGGCGATGC	1185
Db	1180	TTTGGCGCTGTGCTTTTGCCCTATGAAAAGTTTCTGACTTCAGAAAAGCTGTGGCGATGC	1233
OY	1186	CATGAATTTGCTGGGCTACCGCCGACGCTCAGATCTGAACAAACAGCAAAACCTGTGCTCT	1245
Db	1240	CATGAATTTGCTGGGCTACCGCCGACGCTCAGATCTGAACAAACAGCAAAACCTGTGCTCT	1299

QY 1246 GGATCTTCTGTCTACTGAGCTGCTGAGCAATCCACTAAGAGGCTTGAAGAGCTT 1305
 |||||
 Db 1300 GGATCTTCTGTCTACTGAGCTGCTGAGCAATCCACTAAGAGGCTTGAAGAGCTT 1359
 |||||
 QY 1306 TGCTGCCACCTGCTGTGACGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCCTACA 1365
 |||||
 Db 1360 TGCTGCCACCTGCTGTGACGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCCTACA 1419
 |||||
 QY 1366 TGCTGCCACCTGCTGTGACGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCCTACA 1425
 |||||
 Db 1420 TGCTGCCACCTGCTGTGACGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCCTACA 1479
 |||||
 QY 1426 TGCTGCCACCTGCTGTGACGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCCTACA 1485
 |||||
 Db 1480 TGCTGCCACCTGCTGTGACGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCCTACA 1539
 |||||
 QY 1486 TGCTGCCACCTGCTGTGACGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCCTACA 1545
 |||||
 Db 1540 TGCTGCCACCTGCTGTGACGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCCTACA 1599
 |||||
 QY 1546 TGCTGCCACCTGCTGTGACGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCCTACA 1605
 |||||
 Db 1600 TGCTGCCACCTGCTGTGACGCTCAGTCACTTTCTGTGAATGCTTGTGAGCCTTGCCTACA 1659
 |||||
 QY 1606 ATCAGTGGAAATGATCCATAAACCTCCCTGTCCACATCTTGCCTAATGGGGAATGGATCT 1665
 |||||
 Db 1660 ATCAGTGGAAATGATCCATAAACCTCCCTGTCCACATCTTGCCTAATGGGGAATGGATCT 1719
 |||||
 QY 1666 TTCCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1703
 |||||
 Db 1720 TTCCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1757
 |||||

RESULT 11
 AF149783 1333 bp mRNA linear PRI 02-JUL-2001
 LOCUS AF149783 Homo sapiens L-selectin ligand sulfotransferase mRNA, complete cds.
 DEFINITION AF149783
 VERSION AF149783.1 GI:13897503
 KEYWORDS
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 REFERENCE 1 (bases 1 to 1333)
 Yeh,J.C., Hirakawa,N., Petryniak,B., Nakayama,J., Elles,L.G.,
 Raduka,D., Hindsgraul,O., Marth,J.D., Lowe,J.B. and Fukuda,M.
 Novel sulfated lymphocyte homing receptors and their control by a
 Core1 extension beta 1,3-N-acetylglicosaminyltransferase
 Cell 105 (7), 957-969 (2001)

JOURNAL MEDLINE 21332592
 PUBMED 11439191
 REFERENCE 2 (bases 1 to 1333)
 AUTHORS Hirakawa,N. and Fukuda,M.
 TITLE Direct Submission
 JOURNAL Submitted (10-MAY-1999) Glycobiology, The Burnham Institute, 10901
 North Torrey Pines Road, La Jolla, CA 92037, USA
 FEATURES
 SOURCE Location/Qualifiers
 1. 1333
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 93. 1253
 /note="LST"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase"
 /protein_id="AAK48417.1"
 /db_xref="GI:13897504"

TRANSLATION="MLPKRMKLLPLVSVMAIALFFHMYSHNISLSMKADPERNH
 VLIVSSRSSSEFVGLFGQPDVEFLDMEAMHWMFHKSTAMLMIAVRDLRAVF
 LQMSVDAVMKPPROSSLFQWENRALSAPACDIIDODEIIPARHLCSQOP
 FEVEKACRSYSHVLKEVRFNFQSLYPLTKDPSLMHIVHLVRDRAVFRSRERIK

BASE COUNT 286 a 393 c 336 g 318 t
 ORIGIN
 Query Match 61.8%; Score 1262; DB 9; Length 1333;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 GDLMDISRTVMGQHEOKLKEDOPYYVMQVICSOLIEYKTIOSLPKALQERLYLVRY
 EDLAPAPVQOTSRMYEFGLEFLDLMQWNTIRKGMGSHAFHTNRDLNYSQAM
 RWSLPYEVKSHLOKACGDAMMLILRYHRHSRDEQRNLLDLSTWYVEQJH"
 107 AAGCTCTCCACTTCAGCAGCAATGCTACTGCTTAAAAAATGAAAGCTCTGCTGTTTGTG 166
 |||||
 Db 72 AAGCTCTCCACTTCAGCAGCAATGCTACTGCTTAAAAAATGAAAGCTCTGCTGTTTGTG 131
 |||||
 QY 167 GTTCCCGATGGGCACTTGGCTATCTTCCCATGATGACGCAACAATGAGCTCC 226
 |||||
 Db 132 GTTCCCGATGGGCACTTGGCTATCTTCCCATGATGACGCAACAATGAGCTCC 191
 |||||
 QY 227 CTGTCTATGAAGGCGACAGCCGAGGCGATGACAGTGTGTTCTTCTGCGGCTCT 286
 |||||
 Db 192 CTGTCTATGAAGGCGACAGCCGAGGCGATGACAGTGTGTTCTTCTGCGGCTCT 251
 |||||
 QY 287 GGCCTCTTTTGTGTGGGCGACGTTTGTGGGACAGCCAGATGTTTCTACTGATGAG 346
 |||||
 Db 252 GGCCTCTTTTGTGTGGGCGACGTTTGTGGGACAGCCAGATGTTTCTACTGATGAG 311
 |||||
 QY 347 CCCGCTGGCAGCGTGTGATGACCTTCAAGAGACAGCAGCGCTGATGCTGACATGGCT 406
 |||||
 Db 312 CCCGCTGGCAGCGTGTGATGACCTTCAAGAGACAGCAGCGCTGATGCTGACATGGCT 371
 |||||
 QY 407 GTGGGGATCTGATACGGGCGCTTCTTGTGGCAGATGAGCGTCTTGTATGCCATACG 466
 |||||
 Db 372 GTGGGGATCTGATACGGGCGCTTCTTGTGGCAGATGAGCGTCTTGTATGCCATACG 431
 |||||
 QY 467 GAACCTGTCCTCCCGGAGACAGTCCAGCCTCTTTCAGTGGAGAACAGCCGGGCGCTGTGT 526
 |||||
 Db 432 GAACCTGTCCTCCCGGAGAGAGTCCAGCCTCTTTCAGTGGAGAACAGCCGGGCGCTGTGT 491
 |||||
 QY 527 TCTGCACTTGCCTGTGACATCATCCACAAAGTAAATCATCCCGGGGCTACATGCAAG 586
 |||||
 Db 492 TCTGCACTTGCCTGTGACATCATCCACAAAGTAAATCATCCCGGGGCTACATGCAAG 551
 |||||
 QY 587 CTCCTGTGACAGTCAACAGACCCCTTGTGAGTGTGGAGAGGCTGCGCTCTACAGCCAC 646
 |||||
 Db 552 CTCCTGTGACAGTCAACAGACCCCTTGTGAGTGTGGAGAGGCTGCGCTCTACAGCCAC 611
 |||||
 QY 647 GTGCTGTCAAGAGAGTGGCTTCTTCAACCTGCAAGTCCCTCTACCCCTGCTGAAGAC 706
 |||||
 Db 612 GTGCTGTCAAGAGAGTGGCTTCTTCAACCTGCAAGTCCCTCTACCCCTGCTGAAGAC 671
 |||||
 QY 707 CCCCTCCCAACCTGCATATGCTGACCTGGTCCGGGACCCCGGGGCGGTCCGCTGCC 766
 |||||
 Db 672 CCCCTCCCAACCTGCATATGCTGACCTGGTCCGGGACCCCGGGGCGGTCCGCTGCC 731
 |||||
 QY 767 CGAGAACGACAAAAGGAGATCTCATGATGATGACATGCTCATTTGTATGGGCGAGCATGAG 826
 |||||
 Db 732 CGAGAACGACAAAAGGAGATCTCATGATGATGACATGCTCATTTGTATGGGCGAGCATGAG 791
 |||||
 QY 827 CAAAAAATCAAGAGAGAGACCAACCTCATATGATGATGACAGTATCTGCCAAAGCCAG 886
 |||||
 Db 792 CAAAAAATCAAGAGAGAGACCAACCTCATATGATGATGACAGTATCTGCCAAAGCCAG 851
 |||||
 QY 887 CTGAGATCTCAAGAGACCATCAGTCTTGGCCAAAGGCGCTGCGAGAAAGCTTACCTGCTT 946
 |||||
 Db 852 CTGAGATCTCAAGAGACCATCAGTCTTGGCCAAAGGCGCTGCGAGAAAGCTTACCTGCTT 911
 |||||
 QY 947 GTGGCATATGAGGAGCTGGCTGAGGCGCCCTGGGCGCCAGATTCGCCAATATGAAATTC 1006
 |||||
 Db 912 GTGGCATATGAGGAGCTGGCTGAGGCGCCCTGGGCGCCAGATTCGCCAATATGAAATTC 971
 |||||
 QY 1007 GTGGATTTGGAATCTTGGCCCATCTTCAGACTGGGTGCTATACATCACCCGAGGCAAG 1066
 |||||
 Db 972 GTGGATTTGGAATCTTGGCCCATCTTCAGACTGGGTGCTATACATCACCCGAGGCAAG 1031
 |||||

QY 1067 GGCATGGTGACACGCTTTCACACAATGCGAGGATGCCCTTAATGCTCCAGGCT 1126
 DB 1032 GGCATGGTGACACGCTTTCACACAATGCGAGGATGCCCTTAATGCTCCAGGCT 1091
 QY 1127 TGGCGCTGGTCTTGGCCCTATGAAGAAGTTTCTCGACTTCAGAACCTGTGGCATGCC 1186
 DB 1092 TGGCGCTGGTCTTGGCCCTATGAAGAAGTTTCTCGACTTCAGAACCTGTGGCATGCC 1151
 QY 1187 ATGAATTTGCTGGGCTACCGCCAGTCAGATCTGAACAAGAACAGAAACCTGTGGCTG 1246
 DB 1152 ATGAATTTGCTGGGCTACCGCCAGTCAGATCTGAACAAGAACAGAAACCTGTGGCTG 1211
 QY 1247 GATCTTCTGTCTACCTGAGACTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTTT 1306
 DB 1212 GATCTTCTGTCTACCTGAGACTGTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTTT 1271
 Y 1307 GCGGCACCTGTGCTGACGCTCAGTACCTTCTCTGAATGCTTCTGAGCTTGCCTACAT 1366
 DB 1272 GCGGCACCTGTGCTGACGCTCAGTACCTTCTCTGAATGCTTCTGAGCTTGCCTACAT 1331
 QY 1367 CT 1368
 DB 1332 CT 1333
 RESULT 12
 AR310485 AR310485 2065 bp DNA linear PAT 12-JUN-2003
 LOCUS Sequence 26 from patent US 6558935.
 DEFINITION AR310485
 ACCESSION AR310485
 VERSION AR310485.1 GI:31703448
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCES
 1 (bases 1 to 2065)
 Tang,X.T., Corley,N.C., Guegler,K.J., Baughn,M.R., Lal,P., Yue,H.,
 Hillman,J.L. and Azimzal,Y.
 Human transferase proteins
 Patent: US 6558935-A 26 06-MAY-2003;
 Location/Qualifiers
 source 1..2065
 /organism="unknown"
 BASE COUNT 466 a 575 c 491 g 533 t
 RIGIN
 Query Match 56.4%; Score 1153; DB 6; Length 2065;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

DB 452 TGTGGGATCTGATACGGGCGCTCTTGTGCGACATGAGCGCTTGTATGCTACAT 511
 QY 466 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAGAAAGCGGGCCCTGTG 525
 DB 512 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAGAAAGCGGGCCCTGTG 571
 QY 526 TTCTGACCTGCTGTGACATCATCCACAAGATGAATCATCCCGGGCTCACTGACAG 585
 DB 572 TTCTGACCTGCTGTGACATCATCCACAAGATGAATCATCCCGGGCTCACTGACAG 630
 QY 586 GCTTCGTGTGATCAACAGCCCTTTG-AGTGGTGGAAGAGCCCTGCTCTACAGCC 644
 DB 631 GCTTCGTGTGATCAACAGCCCTTTGAGTGGAGAAAGCCCTGCTCTACAGCC 690
 QY 645 AGTGGTGTCAAGAGAGTCCGCTTCTCAACCTGACAGTCCCTCAACCGCTCTGAAG 704
 DB 691 AGTGGTGTCAAGAGAGTCCGCTTCTCAACCTGACAGTCCCTCAACCGCTCTGAAG 750
 QY 705 ACCCTTCCTCAACCTGACATATGTCACCTGTGCGGAGACCCCGGGCCGTTCGTT 764
 DB 751 ACCCTTCCTCAACCTGACATATGTCACCTGTGCGGAGACCCCGGGCCGTTCGTT 810
 QY 765 CCGGAGACGCAAAAGGAGATCTATGTTGACAGTGGCATTTGATGGGACAGT 824
 DB 811 CCGGAGACGCAAAAGGAGATCTATGTTGACAGTGGCATTTGATGGGACAGT 870
 QY 825 AGCAAAACTCAAGAGAGAGACCAACCTACTATGTATGATGACATCTGCAAGCC 884
 DB 871 AGCAAAACTCAAGAGAGAGACCAACCTACTATGTATGATGACATCTGCAAGCC 930
 QY 885 AGCTGAGATCTACAAAGACCATTCAGTCTTGGCCCAAGGCCCTGCAAGAACCTTACCTGC 944
 DB 931 AGCTGAGATCTACAAAGACCATTCAGTCTTGGCCCAAGGCCCTGCAAGAACCTTACCTGC 990
 QY 945 TTGTGGGCTGTAGGAGCTGTGCGAGCCCTGTGGGCCAGACTTCCCAATGTATGAAT 1004
 DB 991 TTGTGGGCTGTAGGAGCTGTGCGAGCCCTGTGGGCCAGACTTCCCAATGTATGAAT 1050
 QY 1005 TCGTGGGATTTGAATTTCTGCCCCATCTTCAGACCTGGGTGCATTAACATCACCGAGCA 1064
 DB 1051 TCGTGGGATTTGAATTTCTGCCCCATCTTCAGACCTGGGTGCATTAACATCACCGAGCA 1110
 QY 1065 AGGGCATGGTGAACACGCTTTCACACAATGCGAGGATGCCCTTAATGTCTCCAGG 1124
 DB 1111 AGGGCATGGTGAACACGCTTTCACACAATGCGAGGATGCCCTTAATGTCTCCAGG 1170
 QY 1125 CTTGGCGCTGTGCTTGGCCCTATGAAGAAGTTTCTGCACTTCAGAAAGCCCTGGCGATG 1184
 DB 1171 CTTGGCGCTGTGCTTGGCCCTATGAAGAAGTTTCTGCACTTCAGAAAGCCCTGGCGATG 1230
 QY 1185 CCATGAATTTGCTGGGCTACCGCCAGTCAGATCTGAACAAGAACAGAGAAACCTGTGCG 1244
 DB 1231 CCATGAATTTGCTGGGCTACCGCCAGTCAGATCTGAACAAGAACAGAGAAACCTGTGCG 1290
 QY 1245 TGGATCTTGTGTACTTGGATGCTGCTGTGAGCAATCCACTAAGAGGTTGAAAGGCT 1304
 DB 1291 TGGATCTTGTGTACTTGGATGCTGCTGTGAGCAATCCACTAAGAGGTTGAAAGGCT 1350
 QY 1305 TTGCTGCACTGGTGTGACGCTCAAGTCACTTCTGCAATGCTTCTGAGACCTTGCCTAC 1364
 DB 1351 TTGCTGCACTGGTGTGACGCTCAAGTCACTTCTGCAATGCTTCTGAGACCTTGCCTAC 1410
 QY 1365 ATCTGAGGCTTAACATACATGTCTGTGGGTATCACACTGAGTGTGATGTGTCACAC 1424
 DB 1411 ATCTGAGGCTTAACATACATGTCTGTGGGTATCACACTGAGTGTGATGTGTCACAC 1470
 QY 1425 GTGCTCAAGCAAGAGACTTTTGTGTCAATGCTTGTGTCTAGAAAACAGACTGGGAAAC 1484
 DB 1471 GTGCTCAAGCAAGAGACTTTTGTGTCAATGCTTGTGTCTAGAAAACAGACTGGGAAAC 1530
 QY 1485 TTATGTAGAGACATATCCACAGAGTGAAGAAAGAGGATTTGCTTCTTCTTCTTGATC 1544
 DB 1531 TTATGTAGAGACATATCCACAGAGTGAAGAAAGAGGATTTGCTTCTTCTTCTTGATC 1590


```

source          1. 877
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
BASE COUNT      183 a      266 c      221 g      203 t      4 others
ORIGIN
Query Match      34.0%; Score 695; DB 6; Length 877;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 106 CAAGCTCTTCACATTCAGCAGATGCTACTGCTAAATAAAGTCTCTGTTCT 165
    |||
Db 47 CAAGGCTTCACATTCAGCAGATGCTACTGCTAAATAAAGTCTCTGTTCT 106
OY 166 GGTTCCTCCAGATGGCCATCTTGCTCTATCTTCACATGATACAGCCACAATCAGCTC 225
    |||
Db 107 GGTTCCTCCAGATGGCCATCTTGCTCTATCTTCACATGATACAGCCACAATCAGCTC 166
OY 226 CCTGTCTATGAAGGACACAGCCGAGCATGACAGCTGCTGTTCTCTCTGGGCTC 285
    |||
Db 167 CCTGTCTATGAAGGACACAGCCGAGCATGACAGCTGCTGTTCTCTCTGGGCTC 226
OY 286 TGAGCTCTCTCTTGTGTGGGAGCTTTTGTGGGAGACCCAGATGTTTCTACCTGATGA 345
    |||
Db 227 TGAGCTCTCTCTTGTGTGGGAGCTTTTGTGGGAGACCCAGATGTTTCTACCTGATGA 286
OY 346 GCCCGCTGGACAGCTGTGATGACCTTCAAGCAGAGACCGCTGTGATGCTGACATGGC 405
    |||
Db 287 GCCCGCTGGACAGCTGTGATGACCTTCAAGCAGAGACCGCTGTGATGCTGACATGGC 346
OY 406 TGTGGGGGATGATACAGGGGCGCTCTTGTGCGACATGAGGCTCTTGTATGCTATCAT 465
    |||
Db 347 TGTGGGGGATGATACAGGGGCGCTCTTGTGCGACATGAGGCTCTTGTATGCTATCAT 406
OY 466 GGAACCTGTGCTCCGAGAGACAGTCCAGCCCTTTCAGTGGGAGAGACAGCCGGGCGCTG 525
    |||
Db 407 GGAACCTGTGCTCCGAGAGACAGTCCAGCCCTTTCAGTGGGAGAGACAGCCGGGCGCTG 466
OY 526 TTGTGCACTGCTGTGACATCATCCACAAGATGAATCATCCCGGGGCTCATGACAG 585
    |||
Db 467 TTGTGCACTGCTGTGACATCATCCACAAGATGAATCATCCCGGGGCTCATGACAG 526
OY 586 GCTCCGTGTGAGCAACAGCCCTTGTAGGAGTGGAGAGGCGCTGCGCTCCATACAGCA 645
    |||
Db 527 GCTCCGTGTGAGCAACAGCCCTTGTAGGAGTGGAGAGGCGCTGCGCTCCATACAGCA 586
OY 646 CGTGTGCTCAAGAGAGTGGCTTCTTCAACTGCAAGTCCCTTACCCGCTGCTGAAGA 705
    |||
Db 587 CGTGTGCTCAAGAGAGTGGCTTCTTCAACTGCAAGTCCCTTACCCGCTGCTGAAGA 646
OY 706 CCCCTCCCTCAACCTGATATCGTGACACCTGATCGGGACCCCGGGCGCTGTTCCGTT 765
    |||
Db 647 CCCCTCCCTCAACCTGATATCGTGACACCTGATCGGGACCCCGGGCGCTGTTCCGTT 706
OY 766 CCGAGAACGACAAAGGAGATTCATGATTGACA 800
    |||
Db 707 CCGAGAACGACAAAGGAGATTCATGATTGACA 741

RESULT 15
AX381256      517 bp      DNA      linear      PAT 18-MAR-2002
LOCUS      AX381256
DEFINITION      Sequence 194 from Patent WO0212280.
ACCESSION      AX381256
VERSION      AX381256.1 GI:19576075
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
REFERENCE      1 Pyle,R.A., Xu,J. and Secrist,H.
AUTHORS

```

```

TITLE      Compositions and methods for the therapy and diagnosis of colon
JOURNAL      Cancer
PATENT      Patent: WO 0212280-A 194 14-FEB-2002;
CORIXA      CORPORATION (US)
FEATURES
SOURCE      1. 517
            /organism="Homo sapiens"
            /mol_type="genomic DNA"
            /db_xref="taxon:9606"
BASE COUNT      121 a      130 c      124 g      142 t
ORIGIN
Query Match      19.7%; Score 403; DB 6; Length 517;
Best Local Similarity 100.0%; Pred. No. 3.5e-228;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1301 GGCTTGGCTGCGACCGTGGTGCAGCTCAGTCACTTCTCTGTAATGCTTGTAGCCCTGC 1360
    |||
Db 13 GGCTTGGCTGCGACCGTGGTGCAGCTCAGTCACTTCTCTGTAATGCTTGTAGCCCTGC 72
OY 1361 CTACATCTCTGAGCCCTTAACATACATGCTGTGGGTATACACATGAGTGTGATGTCC 1420
    |||
Db 73 CTACATCTCTGAGCCCTTAACATACATGCTGTGGGTATACACATGAGTGTGATGTCC 132
OY 1421 ACACGTGCTACACAGAGAGACTTGTGTGCTATGCTGTGTCTAGAAAAACAGACTGGG 1480
    |||
Db 133 ACACGTGCTACACAGAGAGACTTGTGTGCTATGCTGTGTCTAGAAAAACAGACTGGG 192
OY 1481 AACCTTATGAGACGACATCCACAGAGGAAGATGCTCTTCTTCTTCTTCTTCTTCTT 1540
    |||
Db 193 AACCTTATGAGACGACATCCACAGAGGAAGATGCTCTTCTTCTTCTTCTTCTTCTTCTT 252
OY 1541 GATCTTCTGTGTGGGAGACTTTCAGAGACTTGTGGCTGTGAGGCTATTAAGCAGAC 1600
    |||
Db 253 GATCTTCTGTGTGGGAGACTTTCAGAGACTTGTGGCTGTGAGGCTATTAAGCAGAC 312
OY 1601 ACAGTATGAGTGAATGATCCATTAACCTCCCTGTCCACATCTTGCCCAATGGGAATG 1660
    |||
Db 313 ACAGTATGAGTGAATGATCCATTAACCTCCCTGTCCACATCTTGCCCAATGGGAATG 372
OY 1661 GATCTTTCACAAAGAGCTCAGCAGATTTTCCAGAGAGATGC 1703
    |||
Db 373 GATCTTTCACAAAGAGCTCAGCAGATTTTCCAGAGAGATGC 415

```

Search completed: August 14, 2003, 23:39:20
 Job time : 7286 secs

	Query Match	Similarity	Score	ID	Length
Best Local	1597	95.7%	1547	2032	
Matches	1597	Conservative	0	Mismatches	1
			0	Indels	0
			0	Gaps	0
QY	106	CAAGGTCCTCCACTTCAGCACAAATCTACTGCTTAAAAAATGAAAGCTCTGCTGTTTCT	1655		
Db	160	CAAGGTCCTCCACTTCAGCACAAATCTACTGCTTAAAAAATGAAAGCTCTGCTGTTTCT	219		
QY	166	GGTTTCCAGATGGGCATCTGGGCTATCTTCTCCATATGACGCCAACAATCAGCTC	225		
Db	220	GGTTTCCAGATGGGCATCTGGGCTATCTTCTCCATATGACGCCAACAATCAGCTC	279		
QY	226	CCGTGCTTAAGAAGCACAGCCGAGATCAGATGCTGTGTTCTCTTCTGGCCCTC	285		
Db	280	CCGTGCTTAAGAAGCACAGCCGAGATCAGATGCTGTGTTCTCTTCTGGCCCTC	339		
QY	286	TGGCTCTTCTTTTGTGGGGCAGCTTTTGGGCGACCCAGATGTTTTCTTACCTGATGGA	345		

466 GGAACCTGTGCCCCGAGAGAGTCCAGCCTCTTTCAGTGGGAGAAACAGCGGGCCCTGTG 525
|||||
520 GGAACCTGTGCCCCGAGAGAGTCCAGCCTCTTTCAGTGGGAGAAACAGCGGGCCCTGTG 579
|||||
526 TTTCGACACCTGCGCTGTGACATCATCCACAAGATGAATCATCCCGGGGCTCATCTGAG 585
580 TTTCGACACCTGCGCTGTGACATCATCCACAAGATGAATCATCCCGGGGCTCATCTGAG 639
|||||
586 GCTTCGTGTGACATCAAGAGCCTTTGAGTGTGTGGAGAAAGGCGCTGCTTACAGCA 645
640 GCTTCGTGTGACATCAAGAGCCTTTGAGTGTGTGGAGAAAGGCGCTGCTTACAGCA 699
|||||
646 CAGTGTGCTCAAGAGAGTGGCGCTTCTCAACCTGACAGTCCCTTACCGGCTGTGAAGA 705
700 CAGTGTGCTCAAGAGAGTGGCGCTTCTCAACCTGACAGTCCCTTACCGGCTGTGAAGA 759
|||||
706 CCGCTCCCTCAACCTGATATCGTGACCTGCTGGGAGCCCGGGGCGCTGTTCGCTTC 765
760 CCGCTCCCTCAACCTGATATCGTGACCTGCTGGGAGCCCGGGGCGCTGTTCGCTTC 819
|||||
766 CCGAGAGCAGCAAAAGGAGATCATGATTTGACAGTCCGATTTGTGATGGGAGCAGCATGA 825
820 CCGAGAGCAGCAAAAGGAGATCATGATTTGACAGTCCGATTTGTGATGGGAGCAGCATGA 879
|||||
826 GCAAAAACCTCAAGAGAGAGAGCAACCTTACTATGTGATGAGGTCACTCTGCAAAAGCA 885
880 GCAAAAACCTCAAGAGAGAGAGCAACCTTACTATGTGATGAGGTCACTCTGCAAAAGCA 939
|||||
886 GCTGTGATATCTACAAGACCATCCAGTCTTGGCCCAAGGCGCTGAGAGAGCCTACGCT 945
940 GCTGTGATATCTACAAGACCATCCAGTCTTGGCCCAAGGCGCTGAGAGAGCCTACGCT 999
|||||
946 TGTGCGCTATGAGAGACCTGGGTGAGGCCCTGTGGCCAGACCTCCGAATGTGATTT 1005
1000 TGTGCGCTATGAGAGACCTGGGTGAGGCCCTGTGGCCAGACCTCCGAATGTGATTT 1059
|||||
1006 CAGTGTGATGGAATCTTGGCCCATCTTTCAGAGACCTGGGTGATTAATCAACCCAGAGCA 1065
1060 CAGTGTGATGGAATCTTGGCCCATCTTTCAGAGACCTGGGTGATTAATCAACCCAGAGCA 1119
|||||
1066 GGGCATGGTGACCAAGCCTTTCACACAAATGCCAGGAGATGCCCTTAATGTCTCCAGGC 1125
1120 GGGCATGGTGACCAAGCCTTTCACACAAATGCCAGGAGATGCCCTTAATGTCTCCAGGC 1179
|||||
1126 TTTGGCGCTGATTTGGCCATGAAAGTTTCTCGACTTCAGAAACCTGTGGCAGTGC 1185
1180 TTTGGCGCTGATTTGGCCATGAAAGTTTCTCGACTTCAGAAACCTGTGGCAGTGC 1239
|||||
1186 CATGAATTTCTGGGCTACCGCCAGCTCAGATCTGAACAAGAAACAGAAACCTGTGCT 1245
1240 CATGAATTTCTGGGCTACCGCCAGCTCAGATCTGAACAAGAAACAGAAACCTGTGCT 1299
|||||
1246 GATATCTTCTGTACCTGAGCTGCTCGTAGCAAAATCAGTAAAGGGTTGAGAAAGCTT 1305
1300 GATATCTTCTGTACCTGAGCTGCTCGTAGCAAAATCAGTAAAGGGTTGAGAAAGCTT 1359
|||||
1306 TGTGCGCACTGTGTAGAGCTCAGTCACTTCTCTAATGCTTCTAGAGCTTGGCTTACA 1365
1360 TGTGCGCACTGTGTAGAGCTCAGTCACTTCTCTAATGCTTCTAGAGCTTGGCTTACA 1419
|||||
1366 TCTCTGAGCCTTAATCACTATGCTGTGGATATCACTAGATGAGTGTGTGTGCTTACA 1425
1420 TCTCTGAGCCTTAATCACTATGCTGTGGATATCACTAGATGAGTGTGTGTGCTTACA 1479
|||||
1426 TGTCTCAAGCAGAAAGACTTTTGTGTCCATGCTTGTGTGTAAGAAACAGACTGGGAACT 1485
1480 TGTCTCAAGCAGAAAGACTTTTGTGTCCATGCTTGTGTGTAAGAAACAGACTGGGAACT 1539
|||||
1486 TATGTGAGCAGCAATCCCAAGAGTGAAGAGGATTTGCTCTTCTTCTTGTGATCT 1545
1540 TATGTGAGCAGCAATCCCAAGAGTGAAGAGGATTTGCTCTTCTTCTTGTGATCT 1599
|||||

QY 1546 TCTGTGTGGGACAGCTTTCAGAGACTTTGTGGCCTGAGAGCCTATTAAAGCAGACAGT 1605
|||||
Db 1600 TCTGTGTGGGACAGCTTTCAGAGACTTTGTGGCCTGAGAGCCTATTAAAGCAGACAGT 1659
|||||
QY 1606 ATCAGTGAATTTGATTCATTAACCTCCCTGTCCATCTTGGCCAAATGGGAAATGATCT 1665
Db 1660 ATCAGTGAATTTGATTCATTAACCTCCCTGTCCATCTTGGCCAAATGGGAAATGATCT 1719
|||||
QY 1666 TTCACCAAGAGCTCACAGACTTTTCCACAGAGATGC 1703
1720 TTCACCAAGAGCTCACAGACTTTTCCACAGAGATGC 1757
|||||

RESULT 3

US-09-998-598-2595
; Sequence 2595, Application US/0998598
; Patent No. US20020150922A1
; GENERAL INFORMATION:
; APPLICANT: Stolck, John A.
; APPLICANT: Xu, Jianshun
; APPLICANT: Chenault, Ruth A.
; APPLICANT: Mesinger, Madelein Joy
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
; FILE REFERENCE: 210121.561
; CURRENT FILING DATE: 2001-11-16
; NUMBER OF SEQ ID NOS: 2606
; SOFTWARE: Corixa Invention Disclosure Database
; SEQ ID NO 2595
; LENGTH: 505
; TYPE: DNA
; ORGANISM: Homo sapiens
; US-09-998-598-2595

Query Match 19.7%; Score 403; DB 10; Length 505;
Best Local Similarity 100.0%; Pred. No. 8.5e-203;
Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCCTTGTGCGACCTGGTGTGACAGCTTCTCTGATGCTTCTGAGCCTTGC 1360
Db 1 GGCCTTGTGCGACCTGGTGTGACAGCTTCTCTGATGCTTCTGAGCCTTGC 60
|||||
QY 1361 CTACATCTCTAGGCTTAATCACTACATGCTGTGGATACACAGTGTGATGCTGTC 1420
Db 61 CTACATCTCTAGGCTTAATCACTACATGCTGTGGATACACAGTGTGATGCTGTC 120
|||||
QY 1421 ACAGTGTCTCAAGCAGAAAGACTTTGTGTCCATGCTGTGTCTAGAAACAGAGCTGGG 1480
Db 121 ACAGTGTCTCAAGCAGAAAGACTTTGTGTCCATGCTGTGTCTAGAAACAGAGCTGGG 180
|||||
QY 1481 AACCTTATGTGAGCAGACATCCACAGTGAAGAGGATTTGCTCTTCTTCTTCTT 1540
Db 181 AACCTTATGTGAGCAGACATCCACAGTGAAGAGGATTTGCTCTTCTTCTTCTT 240
|||||
QY 1541 GATCTTCTGTCTGGGAGACTTTCAGAGACTTTGGGCTGTGAGGCTTAAAGCAGAC 1600
Db 241 GATCTTCTGTCTGGGAGACTTTCAGAGACTTTGGGCTGTGAGGCTTAAAGCAGAC 300
|||||
QY 1601 ACATATCAGTGAATTTGATCAATAACCTTCCCTGTCCATCTTGGCCAAATGGGAAATG 1660
Db 301 ACATATCAGTGAATTTGATCAATAACCTTCCCTGTCCATCTTGGCCAAATGGGAAATG 360
|||||
QY 1661 GATCTTTCACCAAGAGCTCACAGCACTTTTCCACAGAGATGC 1703
Db 361 GATCTTTCACCAAGAGCTCACAGCACTTTTCCACAGAGATGC 403
|||||

RESULT 4

US-09-919-580-194
; Sequence 194, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:

APPLICANT: Pyle, Ruth
 APPLICANT: Xu, Jiangchun
 APPLICANT: Secrist, Heather
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
 FILE REFERENCE: 210121.552
 CURRENT APPLICATION NUMBER: US/09/919,580
 CURRENT FILING DATE: 2001-07-30
 NUMBER OF SEQ ID NOS: 934
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 194
 LENGTH: 517
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-919-580-194

Query Match 19.7%; Score 403; DB 10; Length 517;
 Best Local Similarity 100.0%; Pred. No. 8.5e-203; Indels 0; Gaps 0;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCCTTGTGCGCCAGCTGCTGAGCCCTCAGTCACTTCTCTGTAATGCTTCTGAGCCTTGC 1360
 DB 13 GGCCTTGTGCGCCAGCTGCTGAGCCCTCAGTCACTTCTCTGTAATGCTTCTGAGCCTTGC 72
 QY 1361 CTAACATCTGTGAGCCCTTAATCTACATGTCTGTGGGTATCACTGAGTGTGCTTCTCC 1420
 DB 73 CTAACATCTGTGAGCCCTTAATCTACATGTCTGTGGGTATCACTGAGTGTGCTTCTCC 132
 QY 1421 ACAGCTGCTCAAGCAGAAAGAGCTTGTGCTCAGTCTTGTCTAGAAAAGACAGCTGGGG 1480
 DB 133 ACAGCTGCTCAAGCAGAAAGAGCTTGTGCTCAGTCTTGTCTAGAAAAGACAGCTGGGG 192
 QY 1481 AACCTTATGTGAGCAGCAGATCCACAGCTGAAAGAGGTATGCTCTTCTTCTTCTTCT 1540
 DB 193 AACCTTATGTGAGCAGCAGATCCACAGCTGAAAGAGGTATGCTCTTCTTCTTCTTCT 252
 QY 1541 GATCTTCTGCTGGGAGACATCTGAGAGCTTGTGCTGAGAGGCTTATTAAGCAGAC 1600
 DB 253 GATCTTCTGCTGGGAGACATCTGAGAGCTTGTGCTGAGAGGCTTATTAAGCAGAC 312
 QY 1601 ACAGTATCAGTGAATTTGATCCATAAAGCTCTGCTCAGATCTTCCCAATGGGGATG 1660
 DB 313 ACAGTATCAGTGAATTTGATCCATAAAGCTCTGCTCAGATCTTCCCAATGGGGATG 372
 QY 1661 GATCTTCTGCTGGGAGACATCTGAGAGCTTGTGCTGAGAGGCTTATTAAGCAGATGC 1703
 DB 373 GATCTTCTGCTGGGAGACATCTGAGAGCTTGTGCTGAGAGGCTTATTAAGCAGATGC 415

RESULT 5
 US-09-919-580-264
 Sequence 264, Application US/09919580
 Patent No. US20020110832A1
 GENERAL INFORMATION:
 APPLICANT: Pyle, Ruth
 APPLICANT: Xu, Jiangchun
 APPLICANT: Secrist, Heather
 TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
 TITLE OF INVENTION: AND DIAGNOSIS OF COLON CANCER
 FILE REFERENCE: 210121.552
 CURRENT APPLICATION NUMBER: US/09/919,580
 CURRENT FILING DATE: 2001-07-30
 NUMBER OF SEQ ID NOS: 934
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 264
 LENGTH: 389
 TYPE: DNA
 ORGANISM: Homo sapiens
 US-09-919-580-264

Query Match 5.6%; Score 114; DB 10; Length 389;
 Best Local Similarity 100.0%; Pred. No. 1.3e-49;
 Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGCCATCTGTATTAATAAATCCCAAAATGCTTGTAGATGTCCTTTTATG 1989
 DB 266 TGCCATCTGTATTAATAAATCCCAAAATGCTTGTAGATGTCCTTTTATG 325
 QY 1990 CTCTTAATTAATAGCAGTAATGCTTATTTATGGATCTTAATAAATAA 2043
 DB 326 CTCTTAATTAATAGCAGTAATGCTTATTTATGGATCTTAATAAATAA 379

RESULT 6
 US-09-927-602-1
 Sequence 1, Application US/09927602
 Patent No. US20020061562A1
 GENERAL INFORMATION:
 APPLICANT: Fukuda, Michiko N.
 APPLICANT: Akama, Tomoya O.
 TITLE OF INVENTION: Methods of Treating Macular Corneal
 FILE REFERENCE: P-LJ 4852
 CURRENT APPLICATION NUMBER: US/09/927,602
 CURRENT FILING DATE: 2001-08-09
 PRIOR APPLICATION NUMBER: US 09/638,211
 NUMBER OF SEQ ID NOS: 38
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 1
 LENGTH: 2544
 TYPE: DNA
 ORGANISM: Homo Sapien
 NAME/KEY: CDS
 LOCATION: (693)...(1877)
 US-09-927-602-1

Query Match 2.9%; Score 59; DB 9; Length 2544;
 Best Local Similarity 100.0%; Pred. No. 1.9e-20;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCGCTCTCTACAGCAGCTGCTCAGAGAGTGCCTTCTCAACCTGCAG 682
 DB 1180 AGGCTGCGCGCTCTCTACAGCAGCTGCTCAGAGAGTGCCTTCTCAACCTGCAG 1238

RESULT 7
 US-09-927-602-38
 Sequence 38, Application US/09927602
 Patent No. US20020061562A1
 GENERAL INFORMATION:
 APPLICANT: Fukuda, Michiko N.
 APPLICANT: Akama, Tomoya O.
 TITLE OF INVENTION: Methods of Treating Macular Corneal
 FILE REFERENCE: P-LJ 4852
 CURRENT APPLICATION NUMBER: US/09/927,602
 CURRENT FILING DATE: 2001-08-09
 PRIOR APPLICATION NUMBER: US 09/638,211
 NUMBER OF SEQ ID NOS: 38
 SOFTWARE: FastSeq for Windows Version 4.0
 SEQ ID NO 38
 LENGTH: 48436
 TYPE: DNA
 ORGANISM: Homo Sapien
 US-09-927-602-38

Query Match 2.9%; Score 59; DB 9; Length 48436;
 Best Local Similarity 100.0%; Pred. No. 1.9e-20;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCGCTCTCTACAGCAGCTGCTCAGAGAGTGCCTTCTCAACCTGCAG 682
 DB 48024 AGGCTGCGCGCTCTCTACAGCAGCTGCTCAGAGAGTGCCTTCTCAACCTGCAG 48082

RESULT 8
US-09-816-825-7
; Sequence 7, Application US/09816825
; Patent No. US20010051370A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CON
; CURRENT APPLICATION NUMBER: US/09/816,825
; CURRENT FILING DATE: 2001-03-22
; PRIOR APPLICATION NUMBER: 09/045,284
; PRIOR FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-816-825-7

Query Match 1.8%; Score 37; DB 9; Length 37;
Best Local Similarity 100.0%; Pred. No. 8,9e-09;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 830 AACTCAAGAGGAGGACCAACCTACTATGTATGATGC 866
DB 1 AACTCAAGAGGAGGAGGACCAACCTACTATGTATGATGC 37

RESULT 9
US-10-007-262-7
; Sequence 7, Application US/10007262
; Publication No. US20020164748A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007,262
; CURRENT FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; PRIOR FILING DATE: EARLIER FILING DATE: 1998-11-12
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 7
; LENGTH: 37
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-007-262-7

Query Match 1.8%; Score 37; DB 13; Length 37;
Best Local Similarity 100.0%; Pred. No. 8,9e-09;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 830 AACTCAAGAGGAGGACCAACCTACTATGTATGATGC 866
DB 1 AACTCAAGAGGAGGAGGACCAACCTACTATGTATGATGC 37

RESULT 10
US-09-867-701-9458
; Sequence 9458, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.

; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE OF INVENTION: AND DIAGNOSIS OF OVARIAN CANCER
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9458
; LENGTH: 167
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9458

Query Match 1.1%; Score 22; DB 10; Length 167;
Best Local Similarity 100.0%; Pred. No. 0.79; 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACGACAGAGGAGGTAGAG 44
DB 3 GAAGACGACAGAGGAGGTAGAG 24

RESULT 11
US-09-867-701-9067
; Sequence 9067, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9067
; LENGTH: 171
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9067

Query Match 1.1%; Score 22; DB 10; Length 171;
Best Local Similarity 100.0%; Pred. No. 0.79; 0; Indels 0; Gaps 0;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACGACAGAGGAGGTAGAG 44
DB 9 GAAGACGACAGAGGAGGTAGAG 30

RESULT 12
US-09-867-701-9185
; Sequence 9185, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Aglate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9185
; LENGTH: 187
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9185

Query Match 1.1%; Score 22; DB 10; Length 187;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACGACAGAGGCTAGAGG 44
DB 32 GAAGACGACAGAGGCTAGAGG 53

RESULT 13
US-09-867-701-9284
; Sequence 9284, Application US/09867701
; Patent No. US20020132237A1
; GENERAL INFORMATION:
; APPLICANT: Agilate, Paul A.
; APPLICANT: Jones, Robert
; APPLICANT: Harlocker, Susan L.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.497
; CURRENT APPLICATION NUMBER: US/09/867,701
; CURRENT FILING DATE: 2001-05-29
; NUMBER OF SEQ ID NOS: 10912
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 9284
; LENGTH: 191
; TYPE: DNA
; ORGANISM: Homo sapien
US-09-867-701-9284

Query Match 1.1%; Score 22; DB 10; Length 191;
Best Local Similarity 100.0%; Pred. No. 0.79;
Matches 22; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 23 GAAGACGACAGAGGCTAGAGG 44
DB 29 GAAGACGACAGAGGCTAGAGG 50

RESULT 14
US-09-960-352-11099/C
; Sequence 11099, Application US/09960352
; Patent No. US20020137139A1
; GENERAL INFORMATION:
; APPLICANT: Warren, Wesley C.
; APPLICANT: Tao, Nengding
; APPLICANT: Byatt, John C.
; APPLICANT: Mathialagan, Nagappan
; TITLE OF INVENTION: NUCLEIC ACID AND OTHER MOLECULES ASSOCIATED WITH LACTATION AND
; FILE REFERENCE: 16511.006/37-21(10298)C
; CURRENT APPLICATION NUMBER: US/09/960,352
; CURRENT FILING DATE: 2001-09-24
; NUMBER OF SEQ ID NOS: 15112
; SEQ ID NO 11099
; LENGTH: 398
; TYPE: DNA
; ORGANISM: Bos taurus
; OTHER INFORMATION: Clone ID: 48-BOVMS1-008-Q1-E1-D8
US-09-960-352-11099

Query Match 1.0%; Score 20; DB 10; Length 398;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 175 GATGGCATTGGCTCTAT 194
DB 240 GATGGCATTGGCTCTAT 221

RESULT 15
US-10-027-632-140660
; Sequence 140660, Application US/10027632

; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide
; FILE REFERENCE: 108827.129
; CURRENT APPLICATION NUMBER: US/10/027,632
; CURRENT FILING DATE: 2002-04-30
; PRIOR APPLICATION NUMBER: US 60/218,006
; PRIOR FILING DATE: 2000-07-12
; PRIOR APPLICATION NUMBER: US 60/198,676
; PRIOR FILING DATE: 2000-04-20
; PRIOR APPLICATION NUMBER: US 60/193,483
; PRIOR FILING DATE: 2000-03-29
; PRIOR APPLICATION NUMBER: US 60/185,218
; PRIOR FILING DATE: 2000-02-24
; PRIOR APPLICATION NUMBER: US 60/167,363
; PRIOR FILING DATE: 1999-11-23
; PRIOR APPLICATION NUMBER: US 60/156,358
; PRIOR FILING DATE: 1999-09-28
; PRIOR APPLICATION NUMBER: US 60/146,002
; PRIOR FILING DATE: 1999-08-09
; NUMBER OF SEQ ID NOS: 325720
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 140660
; LENGTH: 570
; TYPE: DNA
; ORGANISM: Human
US-10-027-632-140660

Query Match 1.0%; Score 20; DB 13; Length 570;
Best Local Similarity 100.0%; Pred. No. 9;
Matches 20; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 700 GAAAGACCCCTCCCTCAACC 719
DB 273 GAAAGACCCCTCCCTCAACC 292

Search completed: August 14, 2003, 21:28:26
Job time: 539 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OK nucleic - nucleic search, using sw model

Run on: August 14, 2003, 17:44:36 ; Search time 144 Seconds
(without alignments)
6262.123 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattccatctgtgtggtta.....tgggactcctaataaaaaaa 2043

Scoring table:

Gapop 60.0, Gapext 60.0

Searched: 569978 seqs, 220691566 residues

Word size: 0

Total number of hits satisfying chosen parameters: 1131652

Minimum DB seq length: 10

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database:

Issued_Patents_NA.*
1: /cgn2-6/ptodata/1/ina/5A.COMB.seq:*
2: /cgn2-6/ptodata/1/ina/5B.COMB.seq:*
3: /cgn2-6/ptodata/1/ina/6A.COMB.seq:*
4: /cgn2-6/ptodata/1/ina/6B.COMB.seq:*
5: /cgn2-6/ptodata/1/ina/PTCUTS.COMB.seq:*
6: /cgn2-6/ptodata/1/ina/Backfile1.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1547	75.7	2032	US-09-045-284A-1	Sequence 1, Appl
2	1547	75.7	2032	US-09-190-911-2	Sequence 2, Appl
3	1153	56.4	2065	US-09-786-240-26	Sequence 26, Appl
4	37	1.8	37	US-09-045-284A-7	Sequence 7, Appl
5	37	1.8	37	US-09-190-911-7	Sequence 7, Appl
6	19	0.9	3792	US-08-992-334-1	Sequence 1, Appl
7	19	0.9	3792	US-08-302-752-1	Sequence 1, Appl
8	19	0.9	5234	US-08-992-334-2	Sequence 2, Appl
9	19	0.9	5234	US-08-302-752-2	Sequence 2, Appl
10	19	0.9	6722	US-08-992-334-3	Sequence 3, Appl
11	19	0.9	6722	US-08-302-752-3	Sequence 3, Appl
12	18	0.9	23	US-08-778-494B-112	Sequence 11, App
13	18	0.9	47	US-09-045-284A-8	Sequence 8, Appl
14	18	0.9	47	US-09-190-911-8	Sequence 8, Appl
15	18	0.9	150	US-09-439-313-320	Sequence 320, App
16	18	0.9	150	US-09-352-616A-320	Sequence 320, App
17	18	0.9	150	US-09-232-149A-320	Sequence 320, App
18	18	0.9	201	US-09-020-956-78	Sequence 78, Appl
19	18	0.9	201	US-09-030-607-78	Sequence 78, Appl
20	18	0.9	201	US-09-439-313-78	Sequence 78, Appl
21	18	0.9	201	US-09-352-616A-78	Sequence 78, Appl
22	18	0.9	201	US-09-232-149A-78	Sequence 78, Appl
23	18	0.9	202	US-09-030-607-180	Sequence 180, App
24	18	0.9	202	US-09-439-313-180	Sequence 180, App
25	18	0.9	202	US-09-352-616A-180	Sequence 180, App
26	18	0.9	202	US-09-232-149A-180	Sequence 180, App
27	18	0.9	536	US-09-020-956-69	Sequence 69, Appl

28	18	0.9	536	US-09-030-607-69	Sequence 69, Appl
29	18	0.9	536	US-09-439-313-69	Sequence 69, Appl
30	18	0.9	536	US-09-352-616A-69	Sequence 69, Appl
31	18	0.9	536	US-09-232-149A-69	Sequence 69, Appl
32	18	0.9	699	US-09-107-532A-1333	Sequence 1333, Ap
33	18	0.9	822	US-09-020-956-29	Sequence 29, Appl
34	18	0.9	822	US-09-030-607-29	Sequence 29, Appl
35	18	0.9	822	US-09-439-313-29	Sequence 29, Appl
36	18	0.9	822	US-09-352-616A-29	Sequence 29, Appl
37	18	0.9	822	US-09-232-149A-29	Sequence 29, Appl
38	18	0.9	1197	US-09-257-179-26	Sequence 26, Appl
39	18	0.9	1368	US-09-118-554-61	Sequence 61, Appl
40	18	0.9	1368	US-09-118-627-61	Sequence 61, Appl
41	18	0.9	1368	US-09-602-877A-61	Sequence 61, Appl
42	18	0.9	1913	US-09-599-360B-40	Sequence 40, Appl
43	18	0.9	1927	US-09-336-536-66	Sequence 66, Appl
44	18	0.9	4351	US-09-634-238-25	Sequence 25, Appl
45	17	0.8	141	US-08-036-555B-161	Sequence 161, App

ALIGNMENTS

RESULT 1					
US-09-045-284A-1					
Sequence 1, Application US/09045284A					
Patent No. 6265192					
GENERAL INFORMATION:					
APPLICANT: Bistrup, Annette					
APPLICANT: Rosen, Steven D.					
TITLE OF INVENTION: GYCOYL SULFOTRANSFERASE-3					
FILE REFERENCE: 6510-107US1					
CURRENT APPLICATION NUMBER: US/09/045,284A					
CURRENT FILING DATE: 1998-03-20					
NUMBER OF SEQ ID NOS: 9					
SOFTWARE: FastSeq for Windows Version 4.0					
SEQ ID NO 1					
LENGTH: 2032					
TYPE: DNA					
ORGANISM: Homo sapiens					
US-09-045-284A-1					
Query Match					
Best Local Similarity 75.7%; Score 1547; DB 3; Length 2032;					
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;					
QY	106	CAAGTCTTCACATTCACACATGCTAATGAAAAATGAAGCTCTGCTGTTCT	165	Sequence 69, Appl	
DB	160	CAAGTCTTCACATTCACACATGCTAATGAAAAATGAAGCTCTGCTGTTCT	219	Sequence 26, Appl	
QY	166	GGTTCCTCCAGATGGCATCTGGCTCTATCTTCACATGTAACACCAACATCAGCTC	225	Sequence 7, Appl	
DB	220	GGTTCCTCCAGATGGCATCTGGCTCTATCTTCACATGTAACACCAACATCAGCTC	279	Sequence 1, Appl	
QY	226	CCTGCTATGAAGCAGACCCGAGCAGTGCCTGTTGTTGTTCTGCTGCTGCTC	285	Sequence 2, Appl	
DB	280	CCTGCTATGAAGCAGACCCGAGCAGTGCCTGTTGTTGTTCTGCTGCTGCTC	339	Sequence 3, Appl	
QY	286	TGGCTCTCTTTTGGGGGACGCTTTTGGGACAGCCAGATGTTTCTACCTGATGA	345	Sequence 11, App	
DB	340	TGGCTCTCTTTTGGGGGACGCTTTTGGGACAGCCAGATGTTTCTACCTGATGA	399	Sequence 8, Appl	
QY	346	GCCCGCTGGCAGCTGTGATGACCTTAAGAGAGCAACCGCTGATGCTGCATGCG	405	Sequence 320, App	
DB	400	GCCCGCTGGCAGCTGTGATGACCTTAAGAGAGCAACCGCTGATGCTGCATGCG	459	Sequence 320, App	
QY	406	TGTGGGATGATGATACGGGCTCTTGTGACATGAGCGCTTGTGATGCTACAT	465	Sequence 78, Appl	
DB	460	TGTGGGATGATGATACGGGCTCTTGTGACATGAGCGCTTGTGATGCTACAT	519	Sequence 78, Appl	
QY	466	GGAACCTGTCCCGGAGACAGTCCAGCTCTTTCAGTGGAGAAACCGGCGCTGTG	525	Sequence 180, App	
DB				Sequence 180, App	

Db 520 GGAACCTGGTCCCGGAGACATGTCAGACCTCTTTCAGTGGGAAACACCGGCGCTGTG 579
QY 526 TTCGTGACCTGCTGTGACATCATCCACAAGATGAATATCCCGGGCTCATGTGAC 585
Db 580 TTCGTGACCTGCTGTGACATCATCCACAAGATGAATATCCCGGGCTCATGTGAC 639
QY 586 GCTCCGTGTGAGCAACAGCCCTTTAGAGTGTGAGAAAGGCGTGGCGCTCATGACCA 645
Db 640 GCTCCGTGTGAGCAACAGCCCTTTAGAGTGTGAGAAAGGCGTGGCGCTCATGACCA 699
QY 646 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACGTCCCTTACCGGCTGCTGAAGA 705
Db 700 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACGTCCCTTACCGGCTGCTGAAGA 759
QY 706 CCCCTCCCTCAACCTGATATGCTGACACCTGTGTCGGGACCCCGGGCGTGTCCGTT 765
Db 760 CCCCTCCCTCAACCTGATATGCTGACACCTGTGTCGGGACCCCGGGCGTGTCCGTT 819
QY 766 CCGAGAACGCAAAAGGAGATCTCATGTGATGACAGTCCGATGTGTGAGGAGCATGA 825
Db 820 CCGAGAACGCAAAAGGAGATCTCATGTGATGACAGTCCGATGTGTGAGGAGCATGA 879
QY 826 GCAAAACTCAAGAGAGAGCAACCCCTACTATGTGATGACAGTCTATGCGCAAGCA 885
Db 880 GCAAAACTCAAGAGAGAGCAACCCCTACTATGTGATGACAGTCTATGCGCAAGCA 939
QY 886 GCTGAGATCTCAAGACATCCAGTCTTGTCCCAAGGCGCTCAGAGAACGCTACTGCT 945
Db 940 GCTGAGATCTCAAGACATCCAGTCTTGTCCCAAGGCGCTCAGAGAACGCTACTGCT 999
QY 946 TGTGCGCTATGAGACCTGGCTGTGAGCCCTGTGGCGCCAGACTTCCGGAATGATGAAT 1005
Db 1000 TGTGCGCTATGAGACCTGGCTGTGAGCCCTGTGGCGCCAGACTTCCGGAATGATGAAT 1059
QY 1006 CGTGGATTTGGAATTTCTTCCCATCTCTGACAGCTGGTGCATATACATCACCAGGCA 1065
Db 1060 CGTGGATTTGGAATTTCTTCCCATCTCTGACAGCTGGTGCATATACATCACCAGGCA 1119
QY 1066 GGGCATGAGTGACACGCTTTCACACAAATGCGAGGAGTCCCTTAATGTCTCCAGGC 1125
Db 1120 GGGCATGAGTGAGACGCTTTCACACAAATGCGAGGAGTCCCTTAATGTCTCCAGGC 1179
QY 1126 TTTGGGCTGTGCTTTTCCCTATGAAAAGTTTCTGCACTTCACAAAAGCCGTGTGCGATGC 1185
Db 1180 TTTGGGCTGTGCTTTTCCCTATGAAAAGTTTCTGCACTTCACAAAAGCCGTGTGCGATGC 1239
QY 1186 CATGAATTTGCTGGGCTACCGCCAGCTAGATCTGAACAAGACAGAAACCTGTGTGCT 1245
Db 1240 CATGAATTTGCTGGGCTACCGCCAGCTAGATCTGAACAAGACAGAAACCTGTGTGCT 1299
QY 1246 GGAATCTTGTCTACCTGACGTGTCCCTGAGCAAAATCCACTAAGAGGGTTGAGAGGCTT 1305
Db 1300 GGAATCTTGTCTACCTGACGTGTCCCTGAGCAAAATCCACTAAGAGGGTTGAGAGGCTT 1359
QY 1306 TGTGTCCACCTGGTGTGACGCTTCAGCTTCTGCAATGCTTCTGAGAGCTTGTGCTTACA 1365
Db 1360 TGTGTCCACCTGGTGTGACGCTTCAGCTTCTGCAATGCTTCTGAGAGCTTGTGCTTACA 1419
QY 1366 TGTGTGAGCTTTAACTACATGTCTGTGGGTATCACACTGAGTGTGATGTGTCACAGC 1425
Db 1420 TGTGTGAGCTTTAACTACATGTCTGTGGGTATCACACTGAGTGTGATGTGTCACAGC 1479
QY 1426 TGTGTCAAGAGAGACTTTTGTGTCCAGTGTGTCTTGAAGAAACAGACAGTGGGAACT 1485
Db 1480 TGTGTCAAGAGAGAGACTTTTGTGTCCAGTGTGTCTTGAAGAAACAGACAGTGGGAACT 1539
QY 1486 TATGTGAGAGAGACATCCACAGCTGAAGAGGTAATGCTCTTCTTCTTCTTCTTATCT 1545
Db 1540 TATGTGAGAGAGACATCCACAGCTGAAGAGGTAATGCTCTTCTTCTTCTTCTTATCT 1599
QY 1546 TCTGTCTGTGAGAGACTTGTGAGAGACTTGTGAGAGCTGAGGCTATTAAGACAGACAGT 1605
Db 1600 TCTGTCTGTGAGAGACTTGTGAGAGACTTGTGAGAGCTGAGGCTATTAAGACAGACAGT 1659

QY 1606 ATCAGTGAATGATCCATAAACCCTCCCTGTGTCACATCTTGGCCAAATGGGAATGATCT 1665
Db 1660 ATCAGTGAATGATCCATAAACCCTCCCTGTGTCACATCTTGGCCAAATGGGAATGATCT 1719
QY 1666 TTTACCAAGAGCTCAACACACATTTTCCACAGAGATGC 1703
Db 1720 TTTACCAAGAGCTCAACACACATTTTCCACAGAGATGC 1757

RESULT 2
US-09-190-911-2
; Sequence 2, Application US/09190911
; Patent No. 6363365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmrich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-09-190-911-2

Query Match 75.7%; Score 1547; DB 4; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAAGGTCTTCCACTTCAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTTTCT 165
Db 160 CAAGGTCTTCCACTTCAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTTTCT 219
QY 166 GGTTCACAGATGGCCATCTTGGCTATCTTCCACATGATAGACACAAATCAAGCTC 225
Db 220 GGTTCACAGATGGCCATCTTGGCTATCTTCCACATGATAGACACAAATCAAGCTC 279
QY 226 CCGTGTATGAAGACAGCCGAGGCGATGACAGTGTGCTGCTTCTGCGGCTC 285
Db 280 CCGTGTATGAAGACAGCCGAGGCGATGACAGTGTGCTGCTTCTGCGGCTC 339
QY 286 TGGCTCTCTTTTGTGGGCGAGCTTTTGGGCGACACCCAGATGTTTCTACCTGATGA 345
Db 340 TGGCTCTCTTTTGTGGGCGAGCTTTTGGGCGACACCCAGATGTTTCTACCTGATGA 399
QY 346 GCGCGGCTGGACAGTGTGATGACCTTCAAGACAGACACCGCTGGATGTGACATGCG 405
Db 400 GCGCGGCTGGACAGTGTGATGACCTTCAAGACAGACACCGCTGGATGTGACATGCG 459
QY 406 TGTGCGGAGTGTATACGGGCGCTTCTTGTGTCAGACATGAGCGCTTTTGATGCTACAT 465
Db 460 TGTGCGGAGTGTATACGGGCGCTTCTTGTGTCAGACATGAGCGCTTTTGATGCTACAT 519
QY 466 GGAACCTGTGCCCGAGACAGTCCAGACCTCTTTCAGTGGAGACAGCCGGGCGCTGTG 525
Db 520 GGAACCTGTGCCCGAGAGACAGTCCAGACCTCTTTCAGTGGAGACAGCCGGGCGCTGTG 579
QY 526 TTTGTGACCTGCTGTGACATCATCCACAAGATGAATATCCCGGGCTCATGTGAC 585
Db 580 TTTGTGACCTGCTGTGACATCATCCACAAGATGAATATCCCGGGCTCATGTGAC 639
QY 586 GCTCTGTGTGAGCAACAGCCCTTTAGAGTGTGAGAAAGGCGTGGCGCTCATGACCA 645
Db 640 GCTCTGTGTGAGCAACAGCCCTTTAGAGTGTGAGAAAGGCGTGGCGCTCATGACCA 699

QY 646 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTAAGA 705
 Db 700 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTAAGA 759
 QY 706 CCCCCTCCCTAACCCTGCATATCGTGACCTGGTCCGGGACCCCGGGCCGTTCGATTC 765
 Db 760 CCCCCTCCCTAACCCTGCATATCGTGACCTGGTCCGGGACCCCGGGCCGTTCGATTC 819
 QY 766 CCGGAGACGCAAAAGGAGATGTCATGATGACAGTCCGATGTTGGGGGACGATGA 825
 Db 820 CCGGAGACGCAAAAGGAGATGTCATGATGACAGTCCGATGTTGGGGGACGATGA 879
 QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTACTATGATGATGATGATGATGATGATGAT 885
 Db 880 GCAAAAACCTCAAGAGAGAGACCAACCTACTATGATGATGATGATGATGATGATGATGAT 939
 QY 886 GCTGGAGATCTACAAAGCATTCCAGTCTTGGCCCAAGCCCTGAGAGACGCTACGCT 945
 Db 940 GCTGGAGATCTACAAAGCATTCCAGTCTTGGCCCAAGCCCTGAGAGACGCTACGCT 999
 QY 946 TGTGCGGTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTCCGGAATGATGAT 1005
 Db 1000 TGTGCGGTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTCCGGAATGATGAT 1059
 QY 1006 CGTGGGATTTGGAATTTTGGCCCATCTTTCAGACCTGGGTGATTAACATCACCCGAGCAA 1065
 Db 1060 CGTGGGATTTGGAATTTTGGCCCATCTTTCAGACCTGGGTGATTAACATCACCCGAGCAA 1119
 QY 1066 GGGGATGGGTGACACGCTTTCACACAAATGCCAGGATGCCCTTAATGCTTCCAGGC 1125
 Db 1120 GGGGATGGGTGACACGCTTTCACACAAATGCCAGGATGCCCTTAATGCTTCCAGGC 1179
 QY 1126 TTTGGCGCTGCTTCTTGGCCATATGAAAGTTTCTGACTTCAGAAAGCCTGTGGCATGC 1185
 Db 1180 TTTGGCGCTGCTTCTTGGCCATATGAAAGTTTCTGACTTCAGAAAGCCTGTGGCATGC 1239
 QY 1186 CATGAATTTCTGCTGGGTACCGCCAGTCTGATGTAACAAGACAGAAACCTGTTCCT 1245
 Db 1240 CATGAATTTCTGCTGGGTACCGCCAGTCTGATGTAACAAGACAGAAACCTGTTCCT 1299
 QY 1246 GGATCTTCTGCTTCTTGGCCATATGAAAGTTTCTGACTTCAGAAAGCCTGTGGCATGC 1305
 Db 1300 GGATCTTCTGCTTCTTGGCCATATGAAAGTTTCTGACTTCAGAAAGCCTGTGGCATGC 1359
 QY 1306 TGTGCGACCTGGGTGACGCTTTCAGTCTTCTGTAATGCTTCTGAGCCCTGCTACA 1365
 Db 1360 TGTGCGACCTGGGTGACGCTTTCAGTCTTCTGTAATGCTTCTGAGCCCTGCTACA 1419
 QY 1366 TCTCTGAGCCTTAACTACATGCTGTGGGTATCATCACTGATGATGATGATGATGATGAT 1425
 Db 1420 TCTCTGAGCCTTAACTACATGCTGTGGGTATCATCACTGATGATGATGATGATGATGAT 1479
 QY 1426 TGTGCAAGCAGAAAGGACTTTTGTGTCCATGCTTGTGTAGAAAACAGACTGGGAACT 1485
 Db 1480 TGTGCAAGCAGAAAGGACTTTTGTGTCCATGCTTGTGTAGAAAACAGACTGGGAACT 1539
 QY 1486 TATGTGAGCAGACATCCACAGATGAAGAGGATTTGCTTCTTCTTCTGATTC 1545
 Db 1540 TATGTGAGCAGACATCCACAGATGAAGAGGATTTGCTTCTTCTTCTGATTC 1599
 QY 1546 TCTGTCTGGGAGACTTTCAGAGACTTGTGGCCCTGAGGACCTATTAAGCAGACAGAT 1605
 Db 1600 TCTGTCTGGGAGACTTTCAGAGACTTGTGGCCCTGAGGACCTATTAAGCAGACAGAT 1659
 QY 1606 ATCAGTGGATTTGATCATTAACCTCCCTGTCCACATCTTGGCCCAATGGGAAATGATTC 1665
 Db 1660 ATCAGTGGATTTGATCATTAACCTCCCTGTCCACATCTTGGCCCAATGGGAAATGATTC 1719
 QY 1666 TTCACCAAGAGGCTCACAGACATTTTCCACAGAGATGC 1703
 Db 1720 TTCACCAAGAGGCTCACAGACATTTTCCACAGAGATGC 1757

RESULT 3
 US-09-786-240-26
 : Sequence 26, Application US/09786240
 : Patent No. 6558935
 : GENERAL INFORMATION:
 : APPLICANT: INCYTE PHARMACEUTICALS, INC.
 : APPLICANT: TANG, Y. Tom
 : APPLICANT: CORLEY, Neil C.
 : APPLICANT: GUEGLER, Karl J.
 : APPLICANT: BAUGHN, Mariah R.
 : APPLICANT: LAL, Preeti
 : APPLICANT: YUE, Henry
 : APPLICANT: HILMAN, Jennifer L.
 : APPLICANT: AZIMZAI, Yalda
 : TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
 : FILE REFERENCE: PF-0592 PCT
 : CURRENT APPLICATION NUMBER: US/09/786, 240
 : PRIOR FILING DATE: 2002-03-12
 : PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,6
 : PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11
 : NUMBER OF SEQ ID NOS: 33
 : SOFTWARE: PERL Program
 : SEQ ID NO 26
 : LENGTH: 2065
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : FEATURE:
 : NAME/KEY: misc-feature
 : OTHER INFORMATION: Incyte ID No. 6558935 2617407CB1
 US-09-786-240-26

Query Match 56.4%; Score 1153; DB 4; Length 2065;
 Best Local Similarity 99.6%; Pred. No. 0;
 Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 106 CAAGCTTCCACTTCAGACCAATGCTACTGCTTAAAAAATGAAGCTCCTGCTTTCT 165
 Db 152 CAAGCTTCCACTTCAGACCAATGCTACTGCTTAAAAAATGAAGCTCCTGCTTTCT 211
 QY 166 GGTTCACCAATGGCCATCTTGGCTATCTTCCATGATGACGACCAACATCACCTC 225
 Db 212 GGTTCACCAATGGCCATCTTGGCTATCTTCCATGATGACGACCAACATCACCTC 271
 QY 226 CCTGTATGAAAGCAGACGCCAGCCATGACAGCTGCTGCTTCTTCTGCGCTC 285
 Db 272 CCTGTATGAAAGCAGACGCCAGCCATGACAGCTGCTGCTTCTTCTGCGCTC 331
 QY 286 TGGCTCTTCTTGGGCGCAGCTTTTGGGACACCCAGAGTCTTCTTCACTGATGA 345
 Db 332 TGGCTCTTCTTGGGCGCAGCTTTTGGGACACCCAGAGTCTTCTTCACTGATGA 391
 QY 346 GCCCGCTGGCAGCTGTGGATGACCTTCAAGCAGACGCGCTGATGCTGCACATGCC 405
 Db 392 GCCCGCTGGCAGCTGTGGATGACCTTCAAGCAGACGCGCTGATGCTGCACATGCC 451
 QY 406 TGTGCGGATCTGATTAAGGCGGCTCTTGTGGACATGAGCTCTTGTATGCTACAT 465
 Db 452 TGTGCGGATCTGATTAAGGCGGCTCTTGTGGACATGAGCTCTTGTATGCTACAT 511
 QY 466 GGAACCTGTGTCGCCGAGACAGTCCACCTCTTCACTGGGAGAAACGCGGCGCTGTG 525
 Db 512 GGAACCTGTGTCGCCGAGAGAGTCCACCTCTTCACTGGGAGAAACGCGGCGCTGTG 571
 QY 526 TTTCTGACCTGCTGTGATCATATCCACAAGATGAATATCCCGCGGCTCACTCAG 585
 Db 572 TTTCTGACCTGCTGTGATCATATCCACAAGATG-AATATCCCGCGGCTCACTCAG 630
 QY 586 GCTCCTGTGAGTCAACAGCCCTTGG-AGGAGTGGGAGAGGCTGCGGCTCAACAGCC 644
 Db 631 GCTCCTGTGAGTCAACAGCCCTTGGAGTGTGGAGAAAGCTGCGCTCAACAGCC 690
 QY 645 ACGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACGCTCCCTTACCCGCTGTAAG 704
 Db 704 ACGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACGCTCCCTTACCCGCTGTAAG

```
Db 691 AGTGTGCTCAAGGAGTGGCTTCTTAACCTGACGTCCCTTAACCCGCTGCTGAAG 750
QY 705 ACCCTCCCTCAACCTGCATATCTGACACCTGGTCCGGAGCCCGCGGCGTTCCTT 764
Db 751 ACCCTCCCTCAACCTGCATATCTGACACCTGGTCCGGAGCCCGCGGCGTTCCTT 810
QY 765 CCCGAGAACGCAAAAGGAGATCTCATATTTGACAGTGGCATTTGATGGGCGACAT 824
Db 811 CCCGAGAACGCAAAAGGAGATCTCATATTTGACAGTGGCATTTGATGGGCGACAT 870
QY 825 AAAAAAATCTAAGAGAGAGACCAACCTACTATGTATGAGGATCATCTGCAAGGC 884
Db 871 AAGCAAAATCTAAGAGAGAGACCAACCTACTATGTATGAGGATCATCTGCAAGGC 930
QY 885 AGCTGAGATCTAACAAGACCATTCAGTCCCTGCGCAAGCCCTGCAGAAAGCTTACT 944
Db 931 AGCTGAGATCTAACAAGACCATTCAGTCCCTGCGCAAGCCCTGCAGAAAGCTTACT 990
QY 945 TTGTGCGCTATAGAGACCTGGCTCGAGCCCTGTGGCCCGACACTTCCGAAATGAT 1004
Db 991 TTGTGCGCTATAGAGACCTGGCTCGAGCCCTGTGGCCCGACACTTCCGAAATGAT 1050
QY 1005 TCGTGGGATTTGAAATTTCTTGCCCATCTTCAAGACCTGGGTGATATACATCACCC 1064
Db 1051 TCGTGGGATTTGAAATTTCTTGCCCATCTTCAAGACCTGGGTGATATACATCACCC 1110
QY 1065 AAGGCAATGGGTGACCAACGCTTTCACACAATGCCAGGAGTCCCTTAATGTCTCC 1124
Db 1111 AAGGCAATGGGTGACCAACGCTTTCACACAATGCCAGGAGTCCCTTAATGTCTCC 1170
QY 1125 CTGGCGCTGTGCTTGGCTTATGAAAAGTTTCTGATTCAGAAAAGCTGTGGGATG 1184
Db 1171 CTGGCGCTGTGCTTGGCTTATGAAAAGTTTCTGATTCAGAAAAGCTGTGGGATG 1230
QY 1185 CCATGAATTTGCTGGGCTACCCGACAGTCAATCTGAACAAGACAGAAACCTGTTC 1244
Db 1231 CCATGAATTTGCTGGGCTACCCGACAGTCAATCTGAACAAGACAGAAACCTGTTC 1290
QY 1245 TGGATCTTCTGCTACCTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAG 1304
Db 1291 TGGATCTTCTGCTACCTGAGCTGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAG 1350
QY 1305 TTGCTGCGACCTGGGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1364
Db 1351 TTGCTGCGACCTGGGTGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1410
QY 1365 ATCTGTGACCTTAATACATGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1424
Db 1411 ATCTGTGACCTTAATACATGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1470
QY 1425 GTGCTCAAGCAGAGAGCTTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1484
Db 1471 GTGCTCAAGCAGAGAGCTTTGTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1530
QY 1485 TATGTGAGCAGACATCCACAGTGAAGAGGATTTGCTGCTGCTGCTGCTGCTGCT 1544
Db 1531 TATGTGAGCAGACATCCACAGTGAAGAGGATTTGCTGCTGCTGCTGCTGCTGCT 1590
QY 1545 TTCTGTCTGGGACAGCTTCAAGACCTTTGGGCTGAGGCTTATTAAGCAGACAG 1604
Db 1591 TTCTGTCTGGGACAGCTTCAAGACCTTTGGGCTGAGGCTTATTAAGCAGACAG 1650
QY 1605 TATCATGTGAATTTGATTAACCTCCCTGTCCACATTTGGCCCAATGGGGAATGAT 1664
Db 1651 TATCATGTGAATTTGATTAACCTCCCTGTCCACATTTGGCCCAATGGGGAATGAT 1710
QY 1665 TTTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1703
Db 1711 TTTCACAAAGAGCTCACAGCATTTTCCACAGAGATGC 1749
```

RESULT 4
US-09-045-284A-7

```
Sequence 7, Application US/09045284A
Patent No. 6265192
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107051
CURRENT APPLICATION NUMBER: US/09/045,284A
CURRENT FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 7
TYPE: DNA
ORGANISM: Homo sapiens
US-09-045-284A-7

Query Match
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 830 AAACCTCAAGAGAGAGGACCAACCTACTATGTGATGC 866
1 AAACCTCAAGAGAGAGGACCAACCTACTATGTGATGC 37

RESULT 5
US-09-190-911-7
Sequence 7, Application US/09190911
Patent No. 6365365
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Fangelmann, Kirsten
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107051
CURRENT APPLICATION NUMBER: US/09/190,911
CURRENT FILING DATE: 1998-11-12
EARLIER APPLICATION NUMBER: 09/045,284
EARLIER FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 7
TYPE: DNA
ORGANISM: H. sapiens
US-09-190-911-7

Query Match
Best Local Similarity 100.0%; Pred. No. 1.2e-08;
Matches 37; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Db 830 AAACCTCAAGAGAGAGGACCAACCTACTATGTGATGC 866
1 AAACCTCAAGAGAGAGGACCAACCTACTATGTGATGC 37

RESULT 6
US-08-992-334-1
Sequence 1, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
TITLE OF INVENTION: PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
```

STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR FR92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: YES
IMMEDIATE SOURCE:
CLONE: pg-host4
US-08-992-334-1

Query Match 0.9%; Score 19; DB 2; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCTGTGACAAAGAGCAAG 1839
1818 TCTGTGACAAAGAGCAAG 1836

RESULT 7
US-08-302-752-1
Sequence 1, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993

INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3792 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-1

Query Match 0.9%; Score 19; DB 3; Length 3792;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1821 TCTGTGACAAAGAGCAAG 1839
1818 TCTGTGACAAAGAGCAAG 1836

RESULT 8
US-08-992-334-2
Sequence 2, Application US/08992334
Patent No. 5916678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christie Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR FR92/03034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-2

Query Match 0.9%; Score 19; DB 2; Length 5234;
Best Local Similarity 100.0%; Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCTGTGACAAAGAGCAAG 1839
DB 3260 TCTGTGACAAAGAGCAAG 3278

RESULT 9

US-08-302-752-2
Sequence 2, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 5234 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-2

Query Match

Best Local Similarity 100.0%, Pred. No. 13;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCTGTGACAAAGAGCAAG 1839
DB 3260 TCTGTGACAAAGAGCAAG 3278

RESULT 10

US-08-992-334-3
Sequence 3, Application US/08992334
Patent No. 5919678
GENERAL INFORMATION:
APPLICANT: Gruss, Alexandra
APPLICANT: Maguin, Emmanuelle
TITLE OF INVENTION: METHODS FOR USING A TEMPERATURE-SENSITIVE
TITLE OF INVENTION: PLASMID
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Christle Parker & Hale, LLP
STREET: 350 West Colorado Boulevard, Suite 500
CITY: Pasadena
STATE: California
COUNTRY: United States
ZIP: 91105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/992,334
FILING DATE: 17-DEC-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:

OY 1821 TCTGTGACAAAGAGCAAG 1839
DB 3260 TCTGTGACAAAGAGCAAG 3278

APPLICATION NUMBER: US 08/302,752
FILING DATE: 24-DEC-1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/FR93/00248
FILING DATE: 12-MAR-1993
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
ATTORNEY/AGENT INFORMATION:
NAME: Prout, D. Bruce
REGISTRATION NUMBER: 20958
REFERENCE/DOCKET NUMBER: C93:31779
TELECOMMUNICATION INFORMATION:
TELEPHONE: (626) 795-9900
TELEFAX: (626) 577-8800
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6722 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
US-08-992-334-3

Query Match

Best Local Similarity 100.0%, Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCTGTGACAAAGAGCAAG 1839
DB 4748 TCTGTGACAAAGAGCAAG 4766

RESULT 11

US-08-302-752-3
Sequence 3, Application US/08302752
Patent No. 6025190
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: THERMOSENSIBLE PLASMID
NUMBER OF SEQUENCES: 3
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/302,752
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: FR 9203034
FILING DATE: 13-MAR-1992
PRIOR APPLICATION DATA:
APPLICATION NUMBER: WO FR/93/00248
FILING DATE: 12-MAR-1993
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 6722 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
US-08-302-752-3

Query Match

Best Local Similarity 100.0%, Pred. No. 14;
Matches 19; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1821 TCTGTGACAAAGAGCAAG 1839
DB 4748 TCTGTGACAAAGAGCAAG 4766

RESULT 12
US-08-778-494B-112
Sequence 112, Application US/08778494B
Patent No. 5962272
GENERAL INFORMATION:
APPLICANT: Chenchik, Alex
APPLICANT: Zhu, York
APPLICANT: Diachenko, Luda
APPLICANT: Siebert, Paul
TITLE OF INVENTION: METHODS AND COMPOSITIONS FOR FULL-LENGTH CDNA
TITLE OF INVENTION: CLONING
NUMBER OF SEQUENCES: 114
CORRESPONDENCE ADDRESS:
ADDRESSEE: Saliwanchik, Lloyd & Saliwanchik
STREET: 2421 N.W. 41st Street, Suite A-1
CITY: Gainesville
STATE: Florida
COUNTRY: USA
ZIP: 32606
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION NUMBER: US/08/778,494B
APPLICATION NUMBER: US/08/778,494B
FILING DATE: 03-JAN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/582,562
FILING DATE: 03-JAN-1996
ATTORNEY/AGENT INFORMATION:
NAME: Pace, Doran R.
REGISTRATION NUMBER: 38,261
REFERENCE/DOCKET NUMBER: CL-7C1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (352) 375-8100
TELEFAX: (352) 372-5800
INFORMATION FOR SEQ ID NO: 112:
SEQUENCE CHARACTERISTICS:
LENGTH: 23 bases
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (synthetic)
US-08-778-494B-112

Query Match 0.9%; Score 18; DB 2; Length 23;
Best Local Similarity 100.0%; Pred. No. 39;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 21 CGGAAGACGACAGAGGG 38
DB 6 CGGAAGACGACAGAGGG 23

RESULT 13
US-09-045-284A-8/C
Sequence 8, Application US/09045284A
Patent No. 6265192
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-1070U1
CURRENT APPLICATION NUMBER: US/09/045,284A
CURRENT FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 8
LENGTH: 47

TYPE: DNA
ORGANISM: Homo sapiens
US-09-045-284A-8

Query Match 0.9%; Score 18; DB 3; Length 47;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1248 ATCTTCGTCTACTCGA 1265
DB 47 ATCTTCGTCTACTCGA 30

RESULT 14
US-09-190-911-8/C
Sequence 8, Application US/09190911
Patent No. 6365365
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Tangemann, Kirsten
APPLICANT: Hemmerich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107CIP
CURRENT APPLICATION NUMBER: US/09/190,911
CURRENT FILING DATE: 1998-11-12
EARLIER APPLICATION NUMBER: 09/045,284
EARLIER FILING DATE: 1998-03-20
NUMBER OF SEQ ID NOS: 8
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 8
LENGTH: 47
TYPE: DNA
ORGANISM: H. sapiens
US-09-190-911-8

Query Match 0.9%; Score 18; DB 4; Length 47;
Best Local Similarity 100.0%; Pred. No. 40;
Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1248 ATCTTCGTCTACTCGA 1265
DB 47 ATCTTCGTCTACTCGA 30

RESULT 15
US-09-439-313-320
Sequence 320, Application US/09439313
Patent No. 6329505
GENERAL INFORMATION:
APPLICANT: Xu, Jiangchun
APPLICANT: Dillon, Davin C.
APPLICANT: Mitcham, Jennifer L.
APPLICANT: Harlocker, Susan Louise
APPLICANT: Jiang Yuqul
APPLICANT: Reed, Steven G.
APPLICANT: Kalos, Michael
APPLICANT: Fanger, Gary
APPLICANT: Reitter, Mark
APPLICANT: Solik, John
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF PROSTATE CANCER
FILE REFERENCE: 210121.427C9
CURRENT APPLICATION NUMBER: US/09/439,313
CURRENT FILING DATE: 1999-11-12
NUMBER OF SEQ ID NOS: 575
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 320
LENGTH: 150
TYPE: DNA
ORGANISM: Homo sapien
US-09-439-313-320

Query Match 0.98; Score 18; DB 4; Length 150;
 Best Local Similarity 100.0%; Pred. No. 41;
 Matches 18; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 GAATTCATTGTGTGGG 18
 ||||||||||||
 Db 31 GAATTCATTGTGTGGG 48

Search completed: August 14, 2003, 21:19:26
 Job time : 157 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 17:40:11 ; Search time 3887 Seconds
(without alignments)
12774.388 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043
Sequence: 1 gaattccattggttggtgta.....tgggatacctaataaaaaaa 2043

Scoring table:
Gapop 60.0 , Gapext 60.0

Searched: 22781392 seqs, 12152238056 residues

Word size : 0

Total number of hits satisfying chosen parameters: 45562670

Minimum DB seq length: 10

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

EST:*
1: em_estda:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_hic:*
9: gb_estl:*
10: gb_est2:*
11: gb_hic:*
12: gb_est3:*
13: gb_est4:*
14: gb_est5:*
15: em_estom:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_huv:*
19: em_gss_pln:*
20: em_gss_vit:*
21: em_gss_fun:*
22: em_gss_man:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vil:*
28: gb_gssl:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	685	33.5	954	12 B1823850 603039012
2	579	28.3	583	2 HSM082755 BX490456 Homo sapi
3	534	26.1	668	9 AL709927 DKF2P5860
4	479	23.4	571	12 B129080 B129080 t1f7c04.y

	c	5	322	15.8	593	12	BM969292
		290	14.2	680	13	BF878439	
		288	14.1	669	13	B0684395	
		218	10.7	553	12	BM128370	
		218	10.7	553	12	BM128370	
		107	5.2	362	9	AM002418	
		105	5.1	417	9	AM572510	
		101	4.9	358	10	BF056840	
		74	3.6	419	9	AM572390	
		14	2.9	525	10	BF197521	
		14	2.9	525	10	BF197521	
		59	2.9	620	9	AI824100	
		59	2.9	695	10	BE857538	
		17	2.9	735	10	BE858652	
		18	2.7	721	28	AO741911	
		19	2.3	436	9	AM081348	
		20	1.3	507	12	BM246681	
		21	1.3	571	12	BM245312	
		22	1.3	695	12	BM245312	
		23	1.3	783	12	BM245312	
		23	1.3	852	12	BM245312	
		24	1.3	852	12	BM245312	
		25	1.3	1923	11	AK009113	
		26	1.1	362	10	BE340900	
		27	1.1	397	9	AI345721	
		28	1.1	687	14	CB167190	
		29	1.1	692	14	BY711808	
		30	1.1	834	13	B0884877	
		31	1.1	148	9	AI345073	
		32	1.1	160	13	B0265877	
		33	1.1	167	9	AI253971	
		34	1.1	167	9	AI253971	
		35	1.1	171	9	AI253971	
		36	1.1	187	9	AI251238	
		37	1.1	187	9	AI251238	
		38	1.1	191	9	AI251979	
		39	1.1	196	9	AI251979	
		40	1.1	253	9	AA076760	
		41	1.1	290	9	AI345395	
		42	1.1	361	9	AI311365	
		43	1.1	493	10	BF869376	
		44	1.1	501	28	BH287165	
		45	1.1	516	9	AA689056	

ALIGNMENTS

RESULT 1
B1823850
LOCUS
DEFINITION
603039012F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179826 5',
mRNA sequence.
B1823850
VERSION
B1823850.1 GI:15935400
KEYWORDS
SOURCE
ORGANISM
Homo sapiens (human)
EST
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
NIH-MGC http://mgc.ncl.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgapbs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1AM11448 row: h column: 03
High quality sequence stop: 856.

QY 281 CGCTTGGCTCTTCTTTTGGGGGAGCTTTTGGGAGACCCAGATGTTTCTACTG 340
 Db 241 CGCTGGGCTCTTCTTTTGGGGGAGCTTTTGGGAGACCCAGATGTTTCTACTG 300
 QY 341 ATGAGAGCCGCTTGGGAGCTTGGGAGATTCATGAGAGAGACCCGCTGGATGCTGAC 400
 Db 301 ATGAGAGCCGCTTGGGAGCTTGGGAGATTCATGAGAGAGACCCGCTGGATGCTGAC 360
 QY 401 ATGAGAGCTTGGGAGATTCATGAGAGAGCCGCTTCTTGGGAGATTCATGAGAGCC 460
 Db 361 ATGAGCTTGGGAGATTCATGAGAGAGCCGCTTCTTGGGAGATTCATGAGAGCC 420
 QY 461 TACATGAGACCTGATCCCGGAGAGACATCCAGCTTTTCAGTGGAGAGACGCGGCGC 520
 Db 421 TACATGAGACCTGATCCCGGAGAGACATCCAGCTTTTCAGTGGAGAGAGCGGCGC 480
 QY 521 CTGTGTTTCTGACCTGCTGCTGATATATCCACAGATGAAATCATCCCGGCTGAC 580
 Db 481 CTGTGTTTCTGACCTGCTGCTGATATATCCACAGATGAAATCATCCCGGCTGAC 540
 QY 581 TGCAGGCTCTGTGACATCAACAGCCCTTGGAGTGTG 619
 Db 541 TGCAGGCTCTGTGACATCAACAGCCCTTGGAGTGTG 579

RESULT 3
 AL709927 668 bp mRNA linear EST 12-JUN-2003
 LOCUS DKFZp68602364.F1 686 (synonym: hlc3) Homo sapiens cDNA clone
 ACCESSION AL709927 GI:19693282
 VERSION AL709927.1 GI:19693282
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 688)
 Authors: Bahr, A., Lauber, J., Mewes, H.W., Well, B., Amid, C., Osanger, A., Fobo, G., Han, M. and Wiemann, S.
 EST (Bahr, A., Lauber, J., Mewes, H.W., Well, B., et al.)
 JOURNAL Unpublished
 COMMENT Contact: Bahr A
 MFS

Ingolstaedter Landstr. 1, D-85764 Neuherberg, Germany
 This is the 5' sequence of the clone insert
 clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by Qiagen (Hilden/Germany) within the cDNA sequencing
 consortium of the German Genome Project.
 No 5' sequence available.
 This clone (DKFZp68602364) is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
 Location/Qualifiers

FEATURES

source
 1. 668
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="DKFZp68602364"
 /tissue_type="human skeletal muscle"
 /dev_stage="adult"
 /lab_host="DH10B"
 /clone_id="686 (synonym: hlc3)"
 /note="vector: pTRIPLEX2; Site_1: SfiIA; Site_2: SfiIB;
 cDNA-collection"
 BASE COUNT 132 a 202 c 178 g 154 t 2 others
 ORIGIN

Query Match 26.1%; Score 534; DB 9; Length 668;
 Best Local Similarity 100.0%; Pred. No. 2.3e-209;
 Matches 534; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 40 AGAGAGAAAAAGCGATGCGCCGCGCTAGCATGAGCTCTCTAAAGAGAGAGAGAGCC 99
 Db 6 AGAGAGAAAAAGCGATGCGCCGCGCTAGCATGAGCTCTCTAAAGAGAGAGAGAGCC 65
 QY 100 AAGCAGAAAGGCTTCCACTTACAGACATGCTACTGCTTAAAAAAATGAAGCTCTGCT 159
 Db 66 AAGCAGAAAGGCTTCCACTTACAGACATGCTACTGCTTAAAAAAATGAAGCTCTGCT 125
 QY 160 GTTTCGTGTTCCAGATGCGCATCTGGCTCTTCTCTTCTCAGATGACAGCCAGCAT 219
 Db 126 GTTTCGTGTTCCAGATGCGCATCTGGCTCTTCTCTTCTCAGATGACAGCCAGCAT 185
 QY 220 CAGCTCCCTCTCTATGAGAGACAGCCGAGCGCATGCTGCTGCTGCTGCTGCTGCT 279
 Db 186 CAGCTCCCTCTCTATGAGAGACAGCCGAGCGCATGCTGCTGCTGCTGCTGCTGCT 245
 QY 280 GCGCTGAGCTCTCTTCTTGGGGAGCTTTTGGGAGACCCAGATGTTTCTACT 339
 Db 246 GCGCTGAGCTCTCTTCTTGGGGAGCTTTTGGGAGACCCAGATGTTTCTACT 305
 QY 340 GATGAGACCGCGCTGGAGAGCTGATGACCTTCAAGCAGAGACCCGCTGGATGCTGCA 399
 Db 306 GATGAGACCGCGCTGGAGAGCTGATGACCTTCAAGCAGAGACCCGCTGGATGCTGCA 365
 QY 400 CATGCTGTGGGAGATCTGATACGGCGCTCTTCTTGGGAGATGAGCGCTTGTGATGC 459
 Db 366 CATGCTGTGGGAGATCTGATACGGCGCTCTTCTTGGGAGATGAGCGCTTGTGATGC 425
 QY 460 CTACATGAAACCTGCTCCCGGAGACAGTCCAGCTCTTTCAGTGGAGAGAACCGGCGC 519
 Db 426 CTACATGAAACCTGCTCCCGGAGACAGTCCAGCTCTTTCAGTGGAGAGAACCGGCGC 485
 QY 520 CCGTGTTCCTGACCTGCTGCTGATCATCATCCCAAGATGAATCATCCCGC 573
 Db 486 CCGTGTTCCTGACCTGCTGCTGATCATCATCCCAAGATGAATCATCCCGC 539

RESULT 4
 BM129080 571 bp mRNA linear EST 12-MAR-2002
 LOCUS I117C04.Y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 DEFINITION cDNA clone IMAGE:567665 5' similar to TR:0915R3
 N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE. ; mRNA sequence.
 ACCESSION BM129080
 VERSION BM129080.1 GI:17123632
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 571)
 Authors: Melton, D., Brown, J., Kenty, G., Permutt, A., Lee, C., Kaestner, K.,
 Lemishka, I., Scarce, M., Brestelli, J., Gradwohl, G., Clifton, S.,
 Haller, L., Marra, M., Pape, D., Wylie, T., Martin, D., Blistain, A.,
 Schmitt, A., Theising, B., Ritter, E., Ronko, I., Bennett, J., Cardenas,
 M., Gibbons, M., McCann, R., Cole, R., Tsagaris, R., Williams, T.,
 Jackson, Y. and Bowers, Y.
 Endocrine Pancreas Consortium
 Unpublished
 Other ESTs: I117C04.x1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute
 Dept. of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@biohp.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Juliana Brown
 (brown@fas.harvard.edu) This sequence now available from the IMAGE
 consortium, for clone orders contact: info@image.llnl.gov

TITLE
 JOURNAL
 COMMENT

FEATURES High quality sequence stop: 434.
Location/Qualifiers
source
1..571
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:567655"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Melton Normalized Human Islet 4 NA-HIS 1"
/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
Site_2: Sal I; Starting library constructed using
Superscript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 Genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 109 a 169 c 143 g 150 t

ORIGIN

Query Match 23.4%; Score 479; DB 12; Length 571;
Best Local Similarity 100.0%; Pred. No. 1.2e-186;
Matches 479; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 CAGGCTTCACCTTCAGCACAATGCTACTGCTTAAATAATGAAGCTCTGCTTTCT 165
|||||
DB 73 CAGGCTTCACCTTCAGCACAATGCTACTGCTTAAATAATGAAGCTCTGCTTTCT 132
|||||
QY 166 GGTTCGCCAGATGGGCATTTGGCTCTATCTTCCACATGTACGCCACAACATAGCTC 225
|||||
DB 133 GGTTCGCCAGATGGGCATTTGGCTCTATCTTCCACATGTACGCCACAACATAGCTC 192
|||||
QY 226 CCTGCTATGAAGGACACACCCGAGGCGATGACGCTGCTTCTTCCGTGGGCGTC 285
|||||
DB 193 CCTGCTATGAAGGACACACCCGAGGCGATGACGCTGCTTCTTCCGTGGGCGTC 252
|||||
QY 286 TGGCTCTTTTGTGGGGGACCTTTTGGGACACCCAGATGTTTCTTACCTGATGA 345
|||||
DB 253 TGGCTCTTTTGTGGGGGACCTTTTGGGACACCCAGATGTTTCTTACCTGATGA 312
|||||
QY 346 GCCGCGCTGACGAGTGTGATGACCTTCAAGACAGACCGGCTGATGCTGACATGGC 405
|||||
DB 313 GCCGCGCTGACGAGTGTGATGACCTTCAAGACAGACCGGCTGATGCTGACATGGC 372
|||||
QY 406 TGTGGGGGATGTGATACGGGCGCTTCTTGTGCGACATGAGCGTCTTGTATGCCAT 465
|||||
DB 373 TGTGGGGGATGTGATACGGGCGCTTCTTGTGCGACATGAGCGTCTTGTATGCCAT 432
|||||
QY 466 GGAACCTGGTCCCCGGAGACATGACAGCTCTTTCAGTGGGAGAACAGCCGGGCGCTGTG 525
|||||
DB 433 GGAACCTGGTCCCCGGAGACATGACAGCTCTTTCAGTGGGAGAACAGCCGGGCGCTGTG 492
|||||
QY 526 TTCTGACACCTGCGCTGACATCATCCACAAGATGAATCAATCCCCGGGCGTACAGCA 584
|||||
DB 493 TTCTGACACCTGCGCTGACATCATCCACAAGATGAATCAATCCCCGGGCGTACAGCA 551
|||||

RESULT 5
BM969292/c 593 bp mRNA linear EST 20-FEB-2003
LOCUS
DEFINITION UI-CF-ENO-2cp-1-21-0-UI s1 UI-CF-ENO Homo sapiens cDNA clone
ACCESSION UI-CF-ENO-2cp-1-21-0-UI 3', mRNA sequence.
BM969292
VERSION BM969292.1 GI:19586879

KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (Dases 1 to 593)
AUTHORS Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE Normalization and subtraction: two approaches to facilitate gene
discovery
JOURNAL Genome Res. 6 (9), 791-806 (1996)
MEDLINE 97044477
PUBMED 8889548
COMMENT Contact: McCray, PB
McCray Lab
University of Iowa
2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
Tel: 319 356 4866
Fax: 319 356 7171
Email: paul-mccray@uiowa.edu
Tissue Procurement: Dr. M. J. Welsh, University of Iowa
cDNA Library preparation: Dr. M. Bento Soares, University of Iowa
cDNA Library Arrayed by: Dr. M. Bento Soares, University of Iowa
DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
Clone Distribution: Researchers may obtain clones from Research
Genetics (www.resgen.com) or from Open Biosystems
(www.openbiosystems.com).
Seq primer: M13 FORWARD
POLYA=Yes.

FEATURES Location/Qualifiers
source
1..593
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="UI-CF-ENO-2cp-1-21-0-UI"
/tissue_type="Primary Lung Cystic Fibrosis Epithelial
Cells"
/dev_stage="Adult"
/lab_host="DH10B (Life Technologies) (T1 phage resistant)"
/clone_lib="UI-CF-ENO"
/note="Organ: Lung; Vector: pT73-Pac (Pharmacia) with a
modified polylinker; Site_1: EcoR I; Site_2: Not I;
UI-CF-ENO is a cDNA library containing the following
tissue(s): Primary Lung Cystic Fibrosis Epithelial Cells.
The library was constructed according to Bonaldo, Lennon
and Soares, Genome Research, 6:791-806, 1996. First strand
cDNA synthesis was primed with an oligo-dT primer
containing a Not I site. Double stranded cDNA was ligated
to an EcoR I adaptor, digested with Not I, and cloned
directionally into pT73-Pac vector. The oligonucleotide
used to prime the synthesis of first-strand cDNA contains
a library tag sequence that is located between the Not I
site and the (dT)18 tail. The sequence tag for this
library is CTGCTCAGT.
TAG LIB=UI-CF-ENO
TAG TISSUE=Human Lung Epithelial Cell Lines untreated LPS
6hr to LPS 24h
TAG_SEQ=CTGCTCAGT"

BASE COUNT 157 a 127 c 137 g 172 t

ORIGIN

Query Match 15.8%; Score 322; DB 12; Length 593;
Best Local Similarity 100.0%; Pred. No. 6.6e-122;
Matches 322; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1382 ACATGCTGTGGTATCACACTGAGTGTGTTGTCACACGTCCTCAGCAGAGA 1441
|||||
DB 593 ACATGCTGTGGTATCACACTGAGTGTGTTGTCACACGTCCTCAGCAGAGA 534
|||||
QY 1442 CTTTGTGTCATGCTTGTGCTAGAAAAACAGACGTGGGACCTTATGTGACAGACAT 1501
|||||
DB 533 CTTTGTGTCATGCTTGTGCTAGAAAAACAGACGTGGGACCTTATGTGACAGACAT 474
|||||
QY 1502 CCCACAGTGAACAGAGGATATGCTCTCTTTCTTGTGATCTTCTGCGGAGAC 1561
|||||

The library was constructed according to Bonaldo, Lennon and Soares, Genome Research, 6:791-806, 1996. First strand cDNA synthesis was primed with an oligo-dT primer containing a Not I site. Double stranded cDNA was ligated to an EcoR I adaptor, digested with Not I, and cloned directionally into p1773-Pac vector. The oligonucleotide used to prime the synthesis of first-strand cDNA contains a library tag sequence that is located between the Not I site and the (dT)18 tail. The sequence tag for this library is CTGCTCAGGT.

TAG_LIB-DI-CF-ENO
TAG_LIB-DI-CF-ENO
TAG_SEQ-CTGCTCAGGT
6hr to LPS 24h

BASE COUNT 191 a 138 c 147 g 193 t

ORIGIN

Query Match 14.1%; Score 288; DB 13; Length 669;
Best Local Similarity 100.0%; Pred. No. 6.7e-108; Mismatches 288; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1416 TGTCCACAGCTGCTCAGACAGAGAGACTTTGTGTCATGCTTGTCTAGAAACAGAC 1475
|||||
636 TGTCCACAGCTGCTCAGACAGAGAGAGACTTTGTGTCATGCTTGTCTAGAAACAGAC 577
|||||
1476 TGGGACACCTTATGTGAGACAGACATCCACAGTGAAGAGGATTCCTCTCTTCT 1535
|||||
576 TGGGACACCTTATGTGAGACAGACATCCACAGTGAAGAGGATTCCTCTCTCTTCT 517
|||||
1536 TCTTGATCTCTGCTGAGAGACTTCAGACATTTGTGCTGAGAGGCTATTAAAC 1595
|||||
516 TTCTTGATCTCTGCTGAGAGACTTCAGACATTTGTGCTGAGAGGCTATTAAAC 457
|||||
1536 ACGACACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGCCCAATGGG 1655
|||||
456 ACGACACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTGCCCAATGGG 397
|||||
1656 GAATGATCTTTCACCAAGAGACTCACCAGCATTTTCCACAGAGATGC 1703
|||||
396 GAATGATCTTTCACCAAGAGACTCACCAGCATTTTCCACAGAGATGC 349
|||||

RESULT 8
BM128370/c 553 bp mRNA linear EST 12-MAR-2002
LOCUS I113d12.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
DEFINITION CDNA clone IMAGE:5676335 3', mRNA sequence.
ACCESSION BM128370
VERSION BM128370.1 GI:17122922
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 553)

TITLE
JOURNAL
COMMENT Endocrine Pancreas Consortium
Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@hchp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For Information on

obtaining a clone please contact: Juliana Brown
(brownjfas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 425.

FEATURES
source

1. 553
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5676335"
/sex="Both"
/tissue_type="Islets of Langerhans"
/dev_stage="Adult"
/lab_host="DH10B"
/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"
/note="Organ: Pancreas; Vector: pSPORFL; Site 1: Not 1;
Site 2: Sal 1; Starting library constructed using
Superscript plasmid library kit (Life Technologies). cDNA
made by oligo-dT priming. Size selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 genome
Research 6:791-806. 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 163 a 111 c 127 g 152 t

ORIGIN

Query Match 10.7%; Score 218; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.3e-79; Mismatches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

1486 TATGAGACAGACATCCACAGTGAAGAGGATTCCTCTCTTCTTGATGCT 1545
|||||
553 TATGAGACAGACATCCACAGTGAAGAGGATTCCTCTCTTCTTGATGCT 494
|||||
1546 TCTGCTGAGAGACTTCAGAGACTTTGTGCTGAGAGGCTATTAAACAGACAGT 1605
|||||
493 TCTGCTGAGAGACTTCAGAGACTTTGTGCTGAGAGGCTATTAAACAGACAGT 434
|||||
1606 ATCAGTGAATTCATTAACCTCCCTGTCACATCTTGCCCAATGGGATGATCT 1665
|||||
433 ATCAGTGAATTCATTAACCTCCCTGTCACATCTTGCCCAATGGGATGATCT 374
|||||
1666 TTCACCAAGAGACTCACCAGCATTTTCCACAGAGATGC 1703
|||||
373 TTCACCAAGAGACTCACCAGCATTTTCCACAGAGATGC 336
|||||

RESULT 9
BM128831/c 553 bp mRNA linear EST 12-MAR-2002
LOCUS I117C04.x1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
DEFINITION CDNA clone IMAGE:5676655 3', mRNA sequence.
ACCESSION BM128831
VERSION BM128831.1 GI:17123383
KEYWORDS EST.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens

REFERENCE
AUTHORS Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (bases 1 to 553)

TITLE
JOURNAL
COMMENT Endocrine Pancreas Consortium
Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@hchp.harvard.edu
Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center For Information on

JOURNAL COMMENT Unpublished
Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
Endocrine Pancreas Consortium
Harvard University, Howard Hughes Medical Institute
Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
MA 02138
Tel: 617-495-1812
Fax: 617-495-8557
Email: dmelton@biochem.harvard.edu

Library was constructed by Dr. Douglas Melton DNA sequencing by:
Washington University Genome Sequencing Center for information on
obtaining a clone please contact: Juliana Brown
(brown@fas.harvard.edu) This sequence now available from the IMAGE
consortium, for clone orders contact: info@image.llnl.gov
High quality sequence stop: 425.

FEATURES
SOURCE

Location/Qualifiers

1..553

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:5676655"

/sex="both"

/tissue_type="Islets of Langerhans"

/dev_stage="Adult"

/lab_host="DH10B"

/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1;
Site_2: Sal 1; Starting library constructed using
SuperScript Plasmid Library kit (Life Technologies). cDNA
made by oligo-dT priming. Size-selected by column
fractionation; average insert size 1.08 kb. Library was
amplified once on solid support and plasmid DNA from
library was prepared. The library DNA was normalized by
method #4 from Bonaldo, Lennon, and Soares 1996 genome
Research 6:791-806; 0.5 microgram single-stranded library
plasmid DNA was mixed with 5 micrograms PCR product
representing library inserts and hybridized to an EcoT of
20. Single-stranded (unhybridized) plasmids were isolated
by hydroxyapatite chromatography and used to make this
library."

BASE COUNT 163 a 111 c 127 g 152 t
ORIGIN

Query Match 10.7%; Score 218; DB 12; Length 553;
Best Local Similarity 100.0%; Pred. No. 5.3e-79;

Matches 218; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1486 TATGTGAGCAGACATCCACCAAGTGAACAGGGTATTGCTCTTCTTTCTTGATCT 1545

Db 553 TATGTGAGCAGACATCCACCAAGTGAACAGGGTATTGCTCTTCTTTCTTGATCT 494

QY 1546 TCCGTCTGGGAGACTTCAGAGACTTGGGCTGAGGCGCTTAAGCAGCAGACAGT 1605

Db 493 TCCGTCTGGGAGACTTCAGAGACTTGGGCTGAGGCGCTTAAGCAGCAGACAGT 434

QY 1606 ATCAGTGAATTGATTCATAAAGCTCCCTGTCCACATCTTGCCCAATGGGAATGATCT 1665

Db 433 ATCAGTGAATTGATTCATAAAGCTCCCTGTCCACATCTTGCCCAATGGGAATGATCT 374

QY 1666 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1703

Db 373 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 336

RESULT 10
AM002418/c 362 bp mRNA linear EST 27-OCT-1999

LOCUS w619g3.x1 NCI_CGAP_G66 Homo sapiens cDNA clone IMAGE:2524564 3',

DEFINITION mRNA sequence.

ACCESSION AM002418

VERSION AM002418.1 GI:5849334

KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM

Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (bases 1 to 362)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

JOURNAL COMMENT

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaps@remail.nih.gov
Tissue Procurement: Christopher A. Moskaluk, M.D., Ph.D., Michael
R. Emmert-Buck, M.D., Ph.D.
CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
Bonaldo, Ph.D.

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington University Genome Sequencing Center
clone distribution: NCI-CGAP clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bdrrp/image/image.html
Seq primer: -40UP from Glibco.

FEATURES

SOURCE

Location/Qualifiers

1..362

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:2524564"

/tissue_type="pooled germ cell tumors"

/lab_host="DH10B"

/clone_id="NCI_CGAP_G66"

/note="Vector: pYT3D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
from the normalized library NCI CGAP GC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (cloneids
1257096-1258631, 1469064-1470983, and 1475592-1476743).
Subtraction by Bento Soares and M. Fatima Bonaldo."

BASE COUNT 108 a 69 c 77 g 108 t
ORIGIN

Query Match 5.2%; Score 107; DB 9; Length 362;
Best Local Similarity 100.0%; Pred. No. 3.7e-33;

Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TGCATCTGTTAAATCTAAATTCCTCAATTAAGTCTCTTTAATAGTCCCTTTATG 1969

Db 107 TGCATCTGTTAAATCTAAATTCCTCAATTAAGTCTCTTTAATAGTCCCTTTATG 48

QY 1990 CTCTTATTTATTTAGAGTAAATGTTCAATTTTATGGAGTCTTAA 2036

Db 47 CTCTTATTTATTTAGAGTAAATGTTCAATTTTATGGAGTCTTAA 1

RESULT 11
AM572510 417 bp mRNA linear EST 13-MAR-2000

LOCUS xq18g11.x2 NCI_CGAP_Ut1 Homo sapiens cDNA clone IMAGE:2751044 3',

DEFINITION mRNA sequence.

ACCESSION AM572510

VERSION AM572510.1 GI:7237243

KEYWORDS EST.
SOURCE Homo sapiens (human)

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 417)
NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
Tumor Gene Index

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaps@remail.nih.gov

JOURNAL COMMENT

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaps@remail.nih.gov

Tissue Procurement: Christopher Koskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 CDNA Library Arrayed by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov/image/html/iresources.shtml
 Seq primer: -40UP from Gibco
 High quality sequence stop: 416.
 Location/Qualifiers

FEATURES

source

```
1. 417
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2751044"
/tissue_type="well-differentiated endometrial
adenocarcinoma, 7 pooled tumors"
/lab_host="DH10B"
/clone_lib="NCI-CGAP-Utl1"
/Note="Organ: uterus; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.75 kb. Life Technologies catalog #:
11538-014"
```

BASE COUNT 140 a 65 c 80 g 132 t

ORIGIN

Query Match 5.1%; Score 105; DB 9; Length 417;
 Best Local Similarity 100.0%; Pred. No. 2.3e-32;
 Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1930 TCCACTCTTTAATCTAAATCCCAATTAAGTCTGTTAGAAATGCCCTTTTATG 1989
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 266 TCCACTCTTTAATCTAAATCCCAATTAAGTCTGTTAGAAATGCCCTTTTATG 207
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1990 CTCTTAATTTATAGCAGTAAATGTTCAATTTTATGCGATCCCTAA 2034
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 206 CTCTTAATTTATAGCAGTAAATGTTCAATTTTATGCGATCCCTAA 162
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 12 358 bp mRNA linear EST 16-OCT-2000
 BF056840/c 7k10d12.x1 NCI-CGAP-CC6 Homo sapiens CDNA clone IMAGE:3443783 3',
 LOCUS mRNA sequence.
 DEINITION BF056840.1 GI:10810736
 ACCESSION EST.
 VERSION Homo sapiens (human)
 KEYWORDS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 SOURCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 ORGANISM National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 COMMENT Unpublished
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 Tissue Procurement: Christopher A. Koskaluk, M.D., Ph.D., Michael
 R. Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D., M. Fatima
 Bonaldi, Ph.D.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL, send email to:
inf@image.llnl.gov
 Seq primer: -40UP from Gibco.

FEATURES

source

```
1. 358
/organism="Homo sapiens"
/mol_type="mRNA"
```

```
/db_xref="taxon:9606"
/clone="IMAGE:3443783"
/tissue_type="pooled germ cell tumors"
/lab_host="DH10B"
/clone_lib="NCI-CGAP-CC6"
/Note="Vector: p773D-Pac (Pharmacia) with a modified
polylinker; Site_1: Not I; Site_2: Eco RI; Plasmid DNA
from the normalized library NCI-CGAP-CC4 was prepared, and
ss circles were made in vitro. Following HAP purification,
this DNA was used as tracer in a subtractive hybridization
reaction. The driver was PCR-amplified cDNAs from a pool
of 5,000 clones made from the same library (clones
1257096-1258631, 1469064-1470983, and 1475592-1476783).
Subtraction by Bento Soares and M. Fatima Bonaldi."
```

BASE COUNT 107 a 68 c 82 g 101 t

ORIGIN

Query Match 4.9%; Score 101; DB 10; Length 358;
 Best Local Similarity 100.0%; Pred. No. 1.1e-30;
 Matches 101; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1932 CCATCTGTTAATCTAAATCCCAATTAAGTCTGTTAGAAATGCCCTTTTATGCT 1991
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 101 CCATCTGTTAATCTAAATCCCAATTAAGTCTGTTAGAAATGCCCTTTTATGCT 42
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1992 TCTTAATTTATAGCAGTAAATGTTCAATTTTATGCGATCCCT 2032
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 41 TCTTAATTTATAGCAGTAAATGTTCAATTTTATGCGATCCCT 1
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 13 419 bp mRNA linear EST 13-MAR-2000
 AW572390/c x009a06.x2 NCI-CGAP-CC14 Homo sapiens CDNA clone IMAGE:2799634 3',
 LOCUS mRNA sequence.
 DEINITION AW572390.1 GI:7237123
 ACCESSION EST.
 VERSION Homo sapiens (human)
 KEYWORDS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 SOURCE NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 ORGANISM National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
 Tumor Gene Index
 COMMENT Unpublished
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgaps-r@mail.nih.gov
 Tissue Procurement: Christopher Koskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Greg Lennon, Ph.D.
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
image.llnl.gov/image/html/iresources.shtml
 Seq primer: -40UP from Gibco
 High quality sequence stop: 343.
 Location/Qualifiers

FEATURES

source

```
1. 419
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:2799634"
/tissue_type="moderately-differentiated adenocarcinoma"
/lab_host="DH10B"
/clone_lib="NCI-CGAP-CC14"
/Note="Organ: colon; Vector: pCMV-SPORT6; Site_1: SalI;
Site_2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Average insert size 1.7 kb. Life Technologies catalog #:
11531-019"
```

BASE COUNT 122 a 78 c 89 g 130 t

ORIGIN

Query Match 3.6%: Score 74; DB 9; Length 419;
 Best Local Similarity 100.0%; Pred. No. 1.4e-19;
 Matches 74; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1630 TCCCTGTCACATCTTGCCCAATGGGAGATGATCTTTCACCAAGAGCTCACCAGCAT 1689
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 419 TCCCTGTCACATCTTGCCCAATGGGAGATGATCTTTCACCAAGAGCTCACCAGCAT 360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 QY 1690 TTCACAGAGATGC 1703
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 359 TTCACAGAGATGC 346
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 14

LOCUS BE197521 525 bp mRNA linear EST 03-NOV-2000
 DEFINITION 708408.x1 NCI-CGAP_Kid11 Homo sapiens cDNA clone IMAGE:3642903 3'

ACCESSION BE197521
 VERSION BE197521
 KEYWORDS 6-O-SULFOTRANSFERASE.; mRNA sequence.
 SOURCE BE197521.1 GI:11086670
 ORGANISM EST.

REFERENCE Homo sapiens (human)
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 COMMENT 1 (bases 1 to 525)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL, send email to:
 info@image.lnl.gov
 High quality sequence stop: 451.
 Location/Qualifiers

1. 525
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:3642903"
 /lab_host="DH10B"
 /clone_1lb="NCI-CGAP_Kid11"
 /note="Organ: kidney; Vector: pT73D-Pac (Pharmacia) with
 a modified polylinker; Site_1: Not I; Site_2: Eco RI;
 Plasmid DNA from the normalized library NCI-CGAP_Kid3
 prepared, and ss circles were made in vitro. Following HAP
 purification, this DNA was used as tracer in a subtractive
 hybridization reaction. The diver was PCR-amplified cDNAs
 from a pool of 5,000 clones made from the same library
 (clones 13223/6-13231/1, 1456007-1456775, and
 1500552-1502855). Subtraction by Bento Soares and M.
 Fatima Bonaldo."
 Fatima Bonaldo.

BASE COUNT 90 a 163 c 187 g 85 t
 ORIGIN

Query Match 2.9%: Score 59; DB 10; Length 525;
 Best Local Similarity 100.0%; Pred. No. 2e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCTCTTACAGCAGTGCTGTCAGAGAGGCGCTTTCACCTGCAG 682
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 260 AGGCTGCGCTCTTACAGCAGTGCTGTCAGAGAGGCGCTTTCACCTGCAG 202
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 15

LOCUS AI824100/c 620 bp mRNA linear EST 21-DEC-1999
 DEFINITION wj46c01.x1 NCI-CGAP_Lu19 Homo sapiens cDNA clone IMAGE:2405856 3'

ACCESSION AI824100
 VERSION AI824100
 KEYWORDS repetitive element.; mRNA sequence.
 SOURCE AI824100.1 GI:5444771
 ORGANISM EST.

REFERENCE Homo sapiens (human)
 AUTHORS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 TITLE Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 JOURNAL NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 COMMENT 1 (bases 1 to 620)
 Unpublished
 Contact: Robert Strausberg, Ph.D.
 Email: cgaps-remail.nih.gov
 Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R.
 Emmert-Buck, M.D., Ph.D.
 CDNA Library Preparation: M. Bento Soares, Ph.D.
 DNA Library Arrayed by: Greg Lennon, Ph.D.
 DNA Sequencing by: Washington University Genome Sequencing Center
 Clone distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LNL at:
 www-bio.lnl.gov/dbp/image/image.html
 Insert Length: 917 Std Error: 0.00
 Seq primer: -40UP from glibco
 High quality sequence stop: 490.
 Location/Qualifiers

1. 620
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /clone="IMAGE:2405856"
 /tissue_type="squamous cell carcinoma, poorly
 differentiated (4 pooled tumors, including primary and
 metastatic)"
 /dev_stage="adult"
 /lab_host="DH10B (phage-resistant)"
 /clone_1lb="NCI-CGAP_Lu19"
 /note="Organ: lung; Vector: pT73D-Pac (Pharmacia) with a
 modified polylinker; 1st strand cDNA was prepared from
 pooled lung tumor tissue, and was then primed with a Not I
 - oligo(dT) primer. Double-stranded cDNA was ligated to
 Eco RI adaptors (Pharmacia), digested with Not I and
 cloned into the Not I and Eco RI sites of the modified
 pT73 vector. Library went through one round of
 normalization. Library constructed by Bento Soares and M.
 Fatima Bonaldo."

BASE COUNT 106 a 195 c 227 g 89 t 3 others
 ORIGIN

Query Match 2.9%: Score 59; DB 9; Length 620;
 Best Local Similarity 100.0%; Pred. No. 1.9e-13;
 Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 624 AGGCTGCGCTCTTACAGCAGTGCTGTCAGAGAGGCGCTTTCACCTGCAG 682
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 DB 314 AGGCTGCGCTCTTACAGCAGTGCTGTCAGAGAGGCGCTTTCACCTGCAG 256
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

Search completed: August 14, 2003, 21:16:41
 Job time : 3900 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:38:30 ; Search time 541 Seconds

(Without alignments)
10193.996 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043
Sequence: 1 gaattccattgtgttgta.....tgggacccaataaaaaaa 2043

Scoring table: OLIGO_NUC
Gapop 60.0 , Gapext 60.0

Searched: 2552756 seqs, 1349719017 residues

Word size : 0

Total number of hits satisfying chosen parameters: 5098080

Minimum DB seq length: 10

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

N_Geneseq_19Jun03:*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1547	75.7	1979	22	AAK94229 Human full-length
2	1547	75.7	2032	20	AA220792 Human glycosyl sul
3	1262	61.8	1333	24	AA516947 Human L-selectin s
4	1153	56.4	2065	21	AA294211 Human transferrase
5	695	34.0	877	22	AAK91803 Human CDNA 5'-end
6	695	34.0	877	22	AAK93921 Human CDNA clone I
7	403	19.7	505	24	ABV89280 Human colon cancer
8	403	19.7	517	24	ABK54724 Human colon cancer

9	306	15.0	2988	21	AAK76156 Human OREF ORF1711
10	114	5.6	389	22	ABK54794 Human colon cancer
11	105	5.1	548	22	AAK92588 Human CDNA 3'-end
12	59	2.9	1647	24	AAD24670 Human drug metabol
13	59	2.9	1694	22	AAD02700 Human glycosyl sul
14	59	2.9	2044	22	AAD02699 Human glycosyl sul
15	59	2.9	2170	22	AAD02698 Human glycosyl sul
16	59	2.9	2544	24	ABN89506 Human corneal N-ac
17	59	2.9	48435	24	ABN89533 Human corneal N-ac
18	59	2.9	160552	22	AAD02697 Human glycosyl sul
19	37	1.8	37	20	AAZ20798 PCR primer for gly
20	30	1.5	30	24	ABK70821 Probe for human ge
21	27	1.3	1926	20	AAZ20793 Mouse glycosyl sul
22	23	1.1	1937	21	AA516948 Murine intestinal-
23	23	1.1	1989	22	AAAD02696 Mouse glycosyl sul
24	22	1.1	22	24	ABK70834 PCR primer for hum
25	22	1.1	22	24	AA516960 Human L-selectin s
26	22	1.1	22	24	AA516962 Human L-selectin s
27	22	1.1	22	24	AA516964 Human L-selectin s
28	22	1.1	167	24	ABL86480 Human ovarian can
29	22	1.1	171	24	ABL86089 Human ovarian can
30	22	1.1	187	24	ABL86207 Human ovarian can
31	22	1.1	191	24	ABL86306 Human ovarian can
32	21	1.0	21	24	ABK70835 PCR primer for hum
33	21	1.0	21	24	AA516961 Human L-selectin s
34	21	1.0	10004	22	ABK70835 Human nervous syst
35	21	1.0	10245	22	ABK70835 Human nervous syst
36	20	1.0	88	21	AAC25980 Human secreted pro
37	20	1.0	354	22	AAK55813 Novel human polynu
38	20	1.0	370	22	AAK55341 Novel human polynu
39	20	1.0	381	22	AAK55341 Novel human polynu
40	20	1.0	398	25	AAK55341 Bovine EST associa
41	20	1.0	3763	19	AAV58194 Human myosin L-cha
42	20	1.0	3763	23	AAK58872 DNA encoding novel
43	20	1.0	5814	23	ABN87855 Human ovary specif
44	20	1.0	6147	25	ACC46453 Human dthp protei
45	20	1.0	9516	22	AAK73555 Human immune/haema

ALIGNMENTS

RESULT 1	AAK94229	standard; CDNA: 1979 BP.
ID	AAK94229	
AC	AAK94229	
XX		
DT	06-NOV-2001	(first entry)
XX		
DE	Human full-length CDNA, SEQ ID NO: 2816.	
XX		
KW	Human; full length CDNA; CDNA synthesis; oligo-capping; ss.	
XX		
OS	Homo sapiens.	
XX		
PN	EP1130094-A2.	
FD	05-SEP-2001.	
XX		
PF	07-JUL-2000; 2000EP-0114089.	
XX		
PR	08-JUL-1999; 99JP-0194486.	
XX		
PR	11-JAN-2000; 2000JP-0118774.	
PR	02-MAY-2000; 2000JP-0183765.	
XX		
PA	(HELI-) HELIX RES INST.	
XX		
PI	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;	
XX	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;	
XX	WPI: 2001-524255/58.	
DR	P-PSDB; AAM93309.	

XX 830 Primers useful for synthesizing full length cDNA clones and their
 PT use in genetic manipulation -
 XX
 PS Claim 8: SEQ ID NO 2816: 1380bp + sequence listing: English.
 XX
 CC The invention relates to primers for synthesizing full length cDNA
 CC clones. 830 cDNA molecules encoding a human protein have been
 CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
 CC molecules have been determined. Primers for synthesizing the full length
 CC cDNA are useful for clarifying the function of the protein encoded by
 CC the cDNA. The full length clones were obtained by construction of full
 CC length enriched cDNA libraries that were synthesised by the oligo-capping
 CC method. The primers enable the production of the full length cDNA easily
 CC without any special methods. The present sequence is a full length
 CC human cDNA of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in CD-ROM format directly from EPO.
 XX
 SQ Sequence 1979 BP: 451 A; 542 C; 474 G; 512 T; 0 other:

Query Match 75.7%; Score 1547; DB 22; Length 1979;
 Best Local Similarity 99.9%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAAAGTCTTCCACTTCAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTCT 165
 DB 47 CAAGGCTCTCCACTTCAGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTGTTCT 106
 QY 166 GGTTCCTCCAGATGGCATCTTGGCTTATCTTCCACATGTACAGCCACAAATCAGCTC 225
 DB 107 GGTTCCTCCAGATGGCATCTTGGCTTATCTTCCACATGTACAGCCACAAATCAGCTC 166
 QY 226 CCTGTCTATGAAGGACACAGCCGAGAGCATGACAGTGTGCTGCTTCTTCCGAGCTC 285
 DB 167 CCGTCTATGAAGGACACAGCCGAGAGCATGACAGTGTGCTGCTTCTTCCGAGCTC 226
 QY 286 TGGCTCTTCTTTTGGGGGACCTTTTGGGACAGCCAGATGTTTCTTACCTGATGA 345
 DB 227 TGGCTCTTCTTTTGGGGGACCTTTTGGGACAGCCAGATGTTTCTTACCTGATGA 286
 QY 346 GCCCCTCGGACAGCTGTGATGACCTTCAAGCAGAGACAGCCGCTGGATCTGACATGAC 405
 DB 287 GCCCCTCGGACAGCTGTGATGACCTTCAAGCAGAGACAGCCGCTGGATCTGACATGAC 346
 QY 406 TGTGCGGGATCTGATACGGGCGCTCTTCTTGTGCGACATGAGCGTCTTGTATGCTACAT 465
 DB 347 TGTGCGGGATCTGATACGGGCGCTCTTCTTGTGCGACATGAGCGTCTTGTATGCTACAT 406
 QY 466 GGAACCTGTGTCCTCCGAGACATCCAGCTTCTTCAAGTGGAGAACAGCCGGCCCTGTG 525
 DB 407 GGAACCTGTGTCCTCCGAGACATCCAGCTTCTTCAAGTGGAGAACAGCCGGCCCTGTG 466
 QY 526 TTTCTACACTGCTGTCGATCATCCACAAATGATTAATCATATCCCGGGGCTCAGTCAG 585
 DB 467 TTTCTACACTGCTGTCGATCATCCACAAATGATTAATCATATCCCGGGGCTCAGTCAG 526
 QY 586 GCTCTGTGACATCAACAGCCCTTGTGAGTGTGGAAGAGGCTGCGCTGCTTACAGCCA 645
 DB 527 GCTCTGTGACATCAACAGCCCTTGTGAGTGTGGAAGAGGCTGCGCTGCTTACAGCCA 586
 QY 646 CGTGTGCTCAAGGAGGTCCTTCTTCAACTGCAAGTCCCTCTACCCGCTGTAAGA 705
 DB 587 CGTGTGCTCAAGGAGGTCCTTCTTCAACTGCAAGTCCCTCTACCCGCTGTAAGA 646
 QY 706 CCCCCTCCCAACCTGCAATATGTCACCTGTCGCGGGAGCCCGGGGCGGTGCGGTTTC 765
 DB 647 CCCCCTCCCAACCTGCAATATGTCACCTGTCGCGGGAGCCCGGGGCGGTGCGGTTTC 706
 QY 766 CCGAAGACGACAAAGGAGATCTCATGTTGACAGTCCATTTGATGGGCGAGCATGA 825
 DB 707 CCGAAGACGACAAAGGAGATCTCATGTTGACAGTCCATTTGATGGGCGAGCATGA 766

QY 826 GCAAAACTCAAGAGAGGAGACCAACCTACTATGTGATGACAGGTATCTGCCAAAGCCA 885
 DB 767 GCAAAACTCAAGAGAGGAGACCAACCTACTATGTGATGACAGGTATCTGCCAAAGCCA 826
 QY 886 GCTGGAGATCTCAAAACCATCCAGTCCCTTGGCCCAAGGCGCTGACAGAAAGGCTACTGCT 945
 DB 827 GCTGGAGATCTCAAAACCATCCAGTCCCTTGGCCCAAGGCGCTGACAGAAAGGCTACTGCT 886
 QY 946 TGTGCGCTATGAGAGACCTGCTGAGACCCCTGTGGCCCAAGCTTCCGAAATGATGAAT 1005
 DB 887 TGTGCGCTATGAGAGACCTGCTGAGACCCCTGTGGCCCAAGCTTCCGAAATGATGAAT 946
 QY 1006 CCGGGATTTGGAATTTTGGCCCATTTTCAAGACCTGGGTCATACATCAGCCAGAGCAA 1065
 DB 947 CCGGGATTTGGAATTTTGGCCCATTTTCAAGACCTGGGTCATACATCAGCCAGAGCAA 1006
 QY 1066 GGGCATGGGTGACACAGCTTTCACACAAATGACAGAGATGCGCTTAATGTCCAGAGC 1125
 DB 1007 GGGCATGGGTGACACAGCTTTCACACAAATGACAGAGATGCGCTTAATGTCCAGAGC 1066
 QY 1126 TTGGCGCTGCTTTTGGCCCTATGAAAAAGTTCCTGACCTTCAGAAAAAGCTGTGGCATGC 1185
 DB 1067 TTGGCGCTGCTTTTGGCCCTATGAAAAAGTTCCTGACCTTCAGAAAAAGCTGTGGCATGC 1126
 QY 1186 CATGAATTTGCTGGGCTACCCGACAGTCAAGATCTGAACAAGAACAGAAACCTGTTGCT 1245
 DB 1127 CATGAATTTGCTGGGCTACCCGACAGTCAAGATCTGAACAAGAACAGAAACCTGTTGCT 1186
 QY 1246 GGAATTTCTGTCTACCTGACATGTCCTCAGCAAAATCCACTAAGAGGTTGAGAAAGCTT 1305
 DB 1187 GGAATTTCTGTCTACCTGACATGTCCTCAGCAAAATCCACTAAGAGGTTGAGAAAGCTT 1246
 QY 1306 TGTGTCACCTGCTGTCACCTTCAGTCACTTCTCTGAATGCTTCTGAGCCCTGCTTACA 1365
 DB 1247 TGTGTCACCTGCTGTCACCTTCAGTCACTTCTCTGAATGCTTCTGAGCCCTGCTTACA 1306
 QY 1366 TCTCTAGGCTTAACATACATGTCGTGGGTATACACTAGTGTAGTGTGTGCCACAG 1425
 DB 1307 TCTCTAGGCTTAACATACATGTCGTGGGTATACACTAGTGTAGTGTGTGCCACAG 1366
 QY 1426 TGTCTAAGCAGAAAGACTTTTGTCTATGCTGTCATGCTTGTCTAAGAAACAGACTGGGAACCT 1485
 DB 1367 TGTCTAAGCAGAAAGACTTTTGTCTATGCTGTCATGCTTGTCTAAGAAACAGACTGGGAACCT 1426
 QY 1486 TATGAGAGCAGACATCCACAGTGAAGAGGATGTCCTCTTCTTCTTCTTCTGATCT 1545
 DB 1427 TATGAGAGCAGACATCCACAGTGAAGAGGATGTCCTCTTCTTCTTCTTCTGATCT 1486
 QY 1546 TCTCTGTGCGGAGATTCAGAGACTTGTGGCTGGAGGCTTATTAACAGACACAGT 1605
 DB 1487 TCTCTGTGCGGAGATTCAGAGACTTGTGGCTGGAGGCTTATTAACAGACACAGT 1546
 QY 1606 ATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTGGCCAAATGGGAATGATCT 1665
 DB 1547 ATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTGGCCAAATGGGAATGATCT 1606
 QY 1666 TTACCAAAAGAGCTACACAGCATTTTCCACAGAGATGC 1703
 DB 1607 TTACCAAAAGAGCTACACAGCATTTTCCACAGAGATGC 1644

RESULT 2
 AA220792
 ID AA220792 standard; DNA; 2032 BP.
 XX
 AC AA220792;
 XX
 DT 08-DEC-1999 (first entry)
 XX
 DE Human glycosyl sulfotransferase-3 coding sequence.
 XX
 KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
 selectin binding interaction; inflammation; lymphocyte homing; human;

XX secondary lymph organ; ss.
XX
OS Homo sapiens.
XX
PN WO9949018-A1.
XX
PD 30-SEP-1999.
XX
PF 26-FEB-1999; 99WO-US04316.
XX
PR 20-MAR-1998; 98US-0045284.
PR 12-NOV-1998; 98US-0190911.
XX
PA (REGC) UNIV CALIFORNIA.
PA (SYNT) SYNTX USA INC.
PI Bistrup A, Rosen SD, Tangemann K, Hemmerich S;
XX
DR WPI; 1999-580442/49.
DR P-PSDB; AAY39918.
XX
PT Human and murine glycosyl sulfotransferase 3 and related
XX polynucleotides
PS Claim 4; Fig 1; 59pp; English.

CC This sequence encodes the human *glycoyl sulfolipase-3* (GSL-3) of
CC the invention. The nucleic acid sequences, probes and primers derived
CC from these, proteins and antibodies are useful in detecting homologues.
CC The sequences, antibodies and methods are useful in the diagnosis and
CC treatment of diseases associated with selectin binding interactions,
CC including conditions associated with or resulting from the homing of
CC leukocytes to sites of inflammation and the normal homing of lymphocytes
CC to secondary lymph organs.

5Q Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other;

Query Match	75.7%;	Score 1547;	DB 20;	Length 2032;
Best Local Similarity	99.9%;	Pred. No. 0;		
Matches 1597; Conservative	0;	Mismatches 1;	Indels 0;	Gaps 0;

OY	106	CAAGGCTTTCCACTTCAGACAAATGGTACTGGCTAAAAAAATGAACCTCTCGTGTTCCT	165
Db	160	CAGGCTCTTCCACTTCAGACAAATGGTACTGGCTAAAAAAATGAACCTCTCGTGTTCCT	219
OY	166	GGTTTCCAGATGAGCCATCTTGGCTCTATTCTTCCACATGTATACAGGCCACAACTCAGCTC	225
Db	220	GGTTTCCAGATGAGCCATCTTGGCTCTATTCTTCCACATGTATACAGGCCACAACTCAGCTC	279
OY	226	CCTGTCTATGAAGGCACAGCCCGAGGGCATGCACGTGCTGTGTCTTCTTCTGGGGCTC	285
Db	280	CCTGTCTATGAAGGCACAGCCCGAGGGCATGCACGTGCTGTGTCTTCTTCTGGGGCTC	339
OY	286	TGGCTCTTCTTTTGTGGGGCAGCTTTTGGGCAGACCCAGATGTTTTCACCTGATATGA	345
Db	340	TGGCTCTTCTTTTGTGGGGCAGCTTTTGGGCAGACCCAGATGTTTTCACCTGATATGA	399
OY	346	GCCCGCTGCGCACGTGTGATGACTTTCAGACGAGACACCGCCTGGATGCTGCATATGCC	405
Db	400	GCCCGCTGCGCACGTGTGATGACTTTCAGACGAGACACCGCCTGGATGCTGCATATGCC	459
OY	406	TGTGGGGGATCTGATACGGGCGCTCTTCTTGTGCAGCAATGAGAGGCTCTTGTATGCCATAT	465
Db	460	TGTGGGGGATCTGATACGGGCGCTCTTCTTGTGTGCAGCAATGAGAGGCTCTTGTATGCCATAT	519
OY	466	GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAGAACAGCCGGGCCCTGTG	525
Db	520	GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTCAGTGGGAGAACAGCCGGGCCCTGTG	579
OY	526	TTTCTGCACCTGGCTGTGACATCATCCCAACAAGATGAATATCATCCCGCGGGCTACATGAC	585
Db	580	TTTCTGCACCTGGCTGTGACATCATCCCAACAAGATGAATATCATCCCGCGGGCTACATGAC	639

QY	586	GCCTCTGAGAGTCAACAGACCCCTTTGAGGTGGTGGAAAGGCTGGCCGTCTCCATACAGCA	645
Db	640	GCTCTCTGACGTCAACAGCCCTTTGAGGTGGTGGAAAGGCTGGCCGTCTCCATACAGCA	699
QY	646	CGTGTGCTCAAGGAGGTGGCGCTTTCTTCAACCTGCAAGTCCCTCTACCCGCTGCTGAAGA	705
Db	700	CGTGTGCTCAAGGAGGTGGCGCTTTCTTCAACCTGCAAGTCCCTCTACCCGCTGCTGAAGA	759
QY	706	CCCCCTCCCTCAACCTGCAATATCTGCACCTGGTCCGGAGCCCCCGGGCCGTGTCCGTTTC	765
Db	760	CCCCCTCCCTCAACCTGCAATATCTGCACCTGGTCCGGAGCCCCCGGGCCGTGTCCGTTTC	819
QY	766	CCGAAAGAGCAAAAGGGAGATCTCATGTTATGCAAGTGGCATTTGATGGGGCAGCATGA	825
Db	820	CCGAAAGAGCAAAAGGGAGATCTCATGTTATGCAAGTGGCATTTGATGGGGCAGCATGA	879
QY	826	GCAAAACCTCAAGAGGAGAGCAACCCCTACTATGTGATGACAGGTCAATCTGGCCAAAGCCA	885
Db	880	GCAGAACTCAAGAGGAGAGCAACCCCTACTATGTGATGACAGGTCAATCTGGCCAAAGCCA	939
QY	886	GCTGAGATCTACAAAGACCATCCAGTCTTGGCCAAAGGCCCTGACAGAAACGTAACCTGCT	945
Db	940	GCTGAGATCTCTCAAGACCATCCAGTCTTGGCCAAAGGCCCTGACAGAAACGTAACCTGCT	999
QY	946	TGTGGGCATAGAGCACTGGGCTCGAGGCCCTGTGGGCCAGACCTGCCGAATGATGAAT	1000
Db	1000	TGTGGGCATAGAGCACTGGGCTCGAGGCCCTGTGGGCCAGACCTGCCGAATGATGAAT	1050
QY	1006	CGTGGGATTTGGAATCTTGCCCCATCTTCAGACCTGGGTGCATATCATCACCAGAGCA	1060
Db	1060	CGTGGGATTTGGAATCTTGCCCCATCTTCAGACCTGGGTGCATATCATCACCAGAGCA	1110
QY	1066	GGGCATGGGTGACACAGCTTTCCACACAAATGCCAGAGATGCCCTTAATGTCTCCAGGC	1120
Db	1120	GGGCATGGGTGACACAGCTTTCCACACAAATGCCAGAGATGCCCTTAATGTCTCCAGGC	1170
QY	1126	TTGGGGCTGGGCTTTGGCCCTATGAAAAGGTTTCTGACATTCAGAAAAGCCGTGGGGCATGC	1180
Db	1180	TTGGGGCTGGGCTTTGGCCCTATGAAAAGGTTTCTGACATTCAGAAAAGCCGTGGGGCATGC	1230
QY	1186	CATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAAACAAAGAACAGAAACCTGTGGCT	1240
Db	1240	CATGAATTTGCTGGGCTACCGCCAGCTCAGATCTGAAACAAAGAACAGAAACCTGTGGCT	1290
QY	1246	GGATCTTCTGTACCTGGAGCTGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAGCTT	1300
Db	1300	GGATCTTCTGTACCTGGAGCTGTCCCTGAGCAAAATCCACTAAGAGGTTGAGAAGCTT	1350
QY	1306	TGCTGCGACACTGGTGCAGACCCACAGCACTTCTCGAATGGCTTCGAGCCCTTGACCTACA	1360
Db	1360	TGCTGCGACACTGGTGCAGACCCACAGCACTTCTCGAATGGCTTCGAGCCCTTGACCTACA	1410
QY	1366	TCTCTGAGCCTTAACATATGCTGTGGGGTATCACACTAGTGTAGTGTGTGTCCACAG	1420
Db	1420	TCTCTGAGCCTTAACATATGCTGTGGGGTATCACACTAGTGTAGTGTGTGTGTGTCCACAG	1470
QY	1426	TGCTCAAGCAGAGGACTTTGTGTGCATGTGTGTCTTAGAAAACAGACTGGGGAACTT	1480
Db	1480	TGCTCAAGCAGAGGACTTTGTGTGCATGTGTGTCTTAGAAAACAGACTGGGGAACTT	1530
QY	1486	TATGTGAGCAGACATCCACACAGTGAACAGAGGTATTCCTCTCTCTCTTTCTTGTATCT	1540
Db	1540	TATGTGAGCAGACATCCACACAGTGAACAGAGGTATTCCTCTCTCTCTTTCTTGTATCT	1590
QY	1546	TCTCTGTGGGAGACTTCAGAGACTTTGTGGCTGAGAGCCATTAATCAACAGACAGT	1600
Db	1600	TCTCTGTGGGAGACTTCAGAGACTTTGTGGCTGAGAGCCATTAATCAACAGACAGT	1650
QY	1606	ATCAGTGAATTTGATTCATTAACCTCCCTGTCCACATCTTGGCCAAATGGGGAATGATCT	1660
Db	1660	ATCAGTGAATTTGATTCATTAACCTCCCTGTCCACATCTTGGCCAAATGGGGAATGATCT	1710
QY	1666	TTTCACCAAAAGAGCTCACCGACATTTTCCACAGAGATGC	1703

DB 1720 TTCACCAAGAGCTCACGACATTTTCACAGAGATGC 1757

RESULT 3

AAS16947
ID AAS16947 standard; cDNA; 1333 BP.

XX AAS16947;

DT 12-MAR-2002 (first entry)

XX Human L-selectin sulfotransferase-2 (LST-2) cDNA.

XX Human; beta1,3gnt; beta1,3-N-acetylglucosaminyltransferase; MECA-79; ss;
XX L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
XX ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
XX allergic contact dermatitis; lymphoma; chronic pneumonia; LST-2;
XX delayed-type hypersensitivity reaction; hyperplastic thymus; antileuc;
XX antiinflammatory; antipsoriatic; antidiabetic; dermatological;
XX antiallergic.

OS Homo sapiens.

PH Key Location/Qualifiers

FT CDS 111..1253

FT /tag= a

PN /product= "Human LST-2"

XX MO200185177-A1.

XX 15-NOV-2001.

XX 10-MAY-2001; 2001WO-US15452.

XX 11-MAY-2000; 2000US-0569320.

XX (BURN-) BURNHAM INST.

XX Fukuda M, Yeh J, Hirakawa N;

XX WPI: 2002-075226/10.

XX P-PSDB; AAU11274.

XX New enzyme, useful for modifying acceptor molecule, comprises an
XX isolated L-selectin sulfotransferase-2 that directs expression of
XX L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or
XX intestinal GLCNAc 6-sulfotransferase

XX Claim 19; Fig 4; 98pp; English.

XX The present invention provides a method of modifying an acceptor molecule
XX by contacting the acceptor with an isolated

XX beta1,3-N-acetylglucosaminyltransferase (beta1,3gnt) or an active
XX fragment, where beta1,3gnt directs expression of a MECA-79 antigen. The
XX invention also provides a method of treating or preventing an

XX L-selectin-mediated condition by reducing the expression or activity of a
XX beta1,3gnt that directs expression of a MECA-79 antigen. This can be done
XX by administering to the subject an oligosaccharide L-selectin antagonist

XX that inhibits binding of L-selectin to a MECA-79 antigen, for example by
XX administering antibody material that specifically binds beta1,3gnt,
XX and/or a beta1,3gnt antisense nucleic acid molecule. L-selectin

XX sulfotransferase-2 (LST-2) also directs MECA-79 antigen expression.
XX Alternatively, the expression or activity of LST-2 or its active
XX fragment can be reduced in combination with reducing the expression or

XX activity of beta1,3gnt. The method is useful for treating L-selectin
XX mediated conditions such as Crohn's disease and ulcerative colitis,
XX inflammatory disorders of the skin such as allergic contact dermatitis,
XX psoriasis and lichen planus, lymphomas, chronic pneumonia, delayed-type

XX hypersensitivity reactions, diabetes and hyperplastic thymus. This
XX sequence represents cDNA encoding human LST-2.

XX Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;

Query Match 61.8%; Score 1262; DB 24; Length 1333;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	107	AAGGTTTCCACTTACGACACATGCTACTGCTCTTAAAAAATGAAGTCTGCTGTTCTG	166
DB	72	AAGGTTTCCACTTACGACACATGCTACTGCTCTTAAAAAATGAAGTCTGCTGTTCTG	131
QY	167	GTTTCCAGATGGCCCTTGGCTGATTTCTTCACATGTACAGCCACAAATCAATCACTCC	226
DB	132	GTTTCCAGATGGCCCTTGGCTGATTTCTTCACATGTACAGCCACAAATCAATCACTCC	191
QY	227	CTGTCTATGAAGGACACAGCCGAGCCATGACAGTGTCTGTCTTCTGAGGCTCT	286
DB	192	CTGTCTATGAAGGACACAGCCGAGCCATGACAGTGTCTGTCTTCTGAGGCTCT	251
QY	287	GCTCTCTTCTTGTGGGAGCTTTTGGGACACCCAGATGTTTCTTACTGATGAG	346
DB	252	GCTCTCTTCTTGTGGGAGCTTTTGGGACACCCAGATGTTTCTTACTGATGAG	311
QY	347	CCGCGCTGGCAGCTGTGATGACCTTCAAGCAGACGCGCGATGCTGACATGGCT	406
DB	312	CCGCGCTGGCAGCTGTGATGACCTTCAAGCAGACGCGCGATGCTGACATGGCT	371
QY	407	GTCGCGGATCTGATACGGGCGCTCTTGTGCGATGAGGCTTTTGTATGCTACATG	466
DB	372	GTCGCGGATCTGATACGGGCGCTCTTGTGCGATGAGGCTTTTGTATGCTACATG	431
QY	467	GAACCTGTGTCCTCCGAGACAGTCCAGCTCTTCACTGGGAGAAACAGCCGCTGT	526
DB	432	GAACCTGTGTCCTCCGAGACAGTCCAGCTCTTCACTGGGAGAAACAGCCGCTGT	491
QY	527	TCTGACCTGCTGTGATCATCCCAAGATGAAATATGATCCCGGCTCATCTACAG	586
DB	492	TCTGACCTGCTGTGATCATCCCAAGATGAAATATGATCCCGGCTCATCTACAG	551
QY	587	CTCTGTGTGACGTCAACAGCCCTTGTGAGTGTGAGAAAGGCTGCCCTCTACAGCC	646
DB	552	CTCTGTGTGACGTCAACAGCCCTTGTGAGTGTGAGAAAGGCTGCCCTCTACAGCC	611
QY	647	GTCGTCCTCAAGAGTGGCTCTTCAACCTGAGCTGCTTACCCGCTGTGAAAGAC	706
DB	612	GTCGTCCTCAAGAGTGGCTCTTCAACCTGAGCTGCTTACCCGCTGTGAAAGAC	671
QY	707	CCCTCCCTCAACCTGATATGTCGACCTGTGCGGAGACCCCGGCTGTTCCTCC	766
DB	672	CCCTCCCTCAACCTGATATGTCGACCTGTGCGGAGACCCCGGCTGTTCCTCC	731
QY	767	CGAGAACGCAAAAGGAGATCTCATGATTGACAGTGCATTGTGATGGGACATGAG	826
DB	732	CGAGAACGCAAAAGGAGATCTCATGATTGACAGTGCATTGTGATGGGACATGAG	791
QY	827	CAAAAACCTAAGAGAGAGACCAACCTCTTGTGATGATGATGATGATGATGATG	886
DB	792	CAAAAACCTAAGAGAGAGACCAACCTCTTGTGATGATGATGATGATGATGATG	851
QY	887	CTGAGATCTACAGACATCCAGTCTTGTGCAAGGCTCTGAGAGACCTTACCTGCTT	946
DB	852	CTGAGATCTACAGACATCCAGTCTTGTGCAAGGCTCTGAGAGACCTTACCTGCTT	911
QY	947	GTCGCTATGAGAGCTGGCTGTGAGCCCTGTGCGCCAGACTTCCGAATGATGATTC	1006
DB	912	GTCGCTATGAGAGCTGGCTGTGAGCCCTGTGCGCCAGACTTCCGAATGATGATTC	971
QY	1007	GTCGCTATGAGATTTCTGCGCATCTTCAGACTGGGTCATATACATCCCGAGGCAAG	1066
DB	972	GTCGCTATGAGATTTCTGCGCATCTTCAGACTGGGTCATATACATCCCGAGGCAAG	1031
QY	1067	GCGATGGTGACAGCTTTCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGCT	1126
DB	1032	GCGATGGTGAGACAGCTTTCACACAAATGCGAGGATGCCCTTAATGTCTCCAGGCT	1091

QY 1127 TGGGCTGCTGCTTGGCCCTATGAAAGGTTCTGACTTCAGACGAAAGCCTGTGGGATGCC 1186
 DB 1092 TGGGCTGCTGCTTGGCCCTATGAAAGGTTCTGACTTCAGACGAAAGCCTGTGGGATGCC 1151
 QY 1187 ATGAATTTGCTGGGCTACCCGACGTCAGATCTGAACAAGAACAAACCTGTGCTG 1246
 DB 1152 ATGAATTTGCTGGGCTACCCGACGTCAGATCTGAACAAGAACAAACCTGTGCTG 1211
 QY 1247 GATCTTCTGCTACCTGAGCTGTCCCTGAGCAATTCACCTAAGAGGGTTGAAGGCTTT 1306
 DB 1212 GATCTTCTGCTACCTGAGCTGTCCCTGAGCAATTCACCTAAGAGGGTTGAAGGCTTT 1271
 QY 1307 GCTGCCACCTGGTGTACGCTCACTCTCTGTAATGCTTGTGACCTTGCTTACAT 1366
 DB 1272 GCTGCCACCTGGTGTACGCTCACTCTCTGTAATGCTTGTGACCTTGCTTACAT 1331
 QY 1367 CT 1368
 DB 1332 CT 1333

RESULT 4

AA294211 standard; cDNA: 2065 BP.

AA294211:

19-JUN-2000 (first entry)

Human transferase TRNSFS-11 cDNA clone 2617407CBI.

Transferrase; TRNSFS-11; human; antitumor; cell proliferation;
 inflammation; gastrointestinal disorder; developmental disorder;
 genetic disorder; neurological disorder; reproductive disorder;
 smooth muscle disorder; immunological disorder; gene therapy;
 diagnosis; N-acetylglicosamine 6-O-sulfotransferase; ss.

Homo sapiens.

Location/Qualifiers

Key 174..1334

FT CDS /tag= a

PN WO200014251-A2.

PD 16-MAR-2000.

PF 09-SEP-1999; 99WO-US20989.

PR 10-SEP-1998; 98US-0150657.

PR 04-NOV-1998; 98US-0186779.

PR 11-MAY-1999; 99US-0133642.

PA (INCY-) INCYTE PHARM INC.

PI Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;

PI Hillman JL, Azimzal Y;

DR WPI: 2000-256996/22.

DR P-PSDB; AAY79219.

XX Human transferase proteins useful for preventing, diagnosing and

XX treating cancers and developmental, gastrointestinal, genetic,

XX immunological, neurological, reproductive and smooth muscle disorders -

XX Claim 9; Page 104-105; 113pp; English.

XX The present sequence is that of cDNA clone 2617407CBI encoding

XX human transferase TRNSFS-11 (see AAY79219), 1 of 15 claimed human

XX transferase proteins of the invention (see AAY79209-23). The clone

XX was isolated from gall bladder cDNA library GBLANOT01. TRNSFS-11

XX is expressed in dermatologic and gastrointestinal tissues,

XX especially those associated with inflammation and cell

CC proliferation. It shows homology to mouse N-acetylglicosamine
 CC 6-O-sulfotransferase. A polynucleotide comprising nucleotides
 CC 264-332 or 1272-1331 of the present sequence can be used as a DNA
 CC probe. The new human transferases and polynucleotides can be used
 CC in the diagnosis, prevention and treatment (including gene therapy
 CC and antisense therapy) of cancer, developmental disorders,
 CC gastrointestinal disorders, genetic disorders, immunological
 CC disorders, neurological disorders, reproductive disorders, and
 CC smooth muscle disorders.

SQ Sequence 2065 BP; 466 A; 575 C; 491 G; 533 T; 0 other;

Query Match 56.4%; Score 1153; DB 21; Length 2065;

Best Local Similarity 99.6%; Pred. NO. 0;

Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

QY 106 CAAGCTCTTCCACTTCACGACACATGCTCTGCTTAAAAAATGAAGCTCTGCTTCT 165
 DB 152 CAAGCTCTTCCACTTCACGACACATGCTCTGCTTAAAAAATGAAGCTCTGCTTCT 211
 QY 166 GATTCCGAGATGGCCATCTTGGCTCTATCTTCACATGTACAGCCCAACATCAGCTC 225
 DB 212 GATTCCGAGATGGCCATCTTGGCTCTATCTTCACATGTACAGCCCAACATCAGCTC 271
 QY 226 CCTGTCTATGAGGCACAGCCGACGATGCTGCTGCTCTCTCTGCTGCTC 285
 DB 272 CCGTCTATGAGGCACAGCCGACGATGCTGCTGCTCTCTCTGCTGCTC 331
 QY 286 TGGCTCTCTTCTTGTGGGGCAGCTTTTGGGAGCAGCCAGATGTTTCTTACCTATGCA 345
 DB 332 TGGCTCTCTTCTTGTGGGGCAGCTTTTGGGAGCAGCCAGATGTTTCTTACCTATGCA 391
 QY 346 GCCGCCGAGGACGCTGATGAGCTTCAAGAGAGCAGCCGCTGATGCTGACATGAC 405
 DB 392 GCCGCCGAGGACGCTGATGAGCTTCAAGAGAGCAGCCGCTGATGCTGACATGAC 451
 QY 406 TGTGGGGATGTGATACGGGCGCTTCTTGTGACATGAGCGTCTTGTATGCTTACAT 465
 DB 452 TGTGGGGATGTGATACGGGCGCTTCTTGTGACATGAGCGTCTTGTATGCTTACAT 511
 QY 466 GGAACCTGTGCTCCGAGACAGCTCAGCTCTTTAGTGGGAGAACAGCCGGGCTGTG 525
 DB 512 GGAACCTGTGCTCCGAGACAGCTCAGCTCTTTAGTGGGAGAACAGCCGGGCTGTG 571
 QY 526 TTTGCGACCTGCTGTGACATCATCCCAAGATGAATCCCGGGGCTCACTGAG 585
 DB 572 TTTGCGACCTGCTGTGACATCATCCCAAGATGAATCCCGGGGCTCACTGAG 630
 QY 586 GCTCCTGTGACATCAAGCCCTTTG-AGTGTGTGAGAGAGCCGCTCTCTACAGCC 644
 DB 631 GCTCCTGTGACATCAAGCCCTTTGAGTGTGTGAGAGAGCCGCTCTCTACAGCC 690
 QY 645 AGTGTGTGCTCAAGAGGTGCGCTTCTCAACCTGACAGTCCCTTACCCGCTGTAAG 704
 DB 691 AGTGTGTGCTCAAGAGGTGCGCTTCTCAACCTGACAGTCCCTTACCCGCTGTAAG 750
 QY 705 ACCCTGCTCAACCTGATATCCGACACCTGCTGCGGAGACCCCGGGGCTGTGCTT 764
 DB 751 ACCCTGCTCAACCTGATATCCGACACCTGCTGCGGAGACCCCGGGGCTGTGCTT 810
 QY 765 CCCGAGACGACAAAGGAGATCTCATGATTGACAGTGTGATGAGGAGCAGATG 824
 DB 811 CCCGAGACGACAAAGGAGATCTCATGATTGACAGTGTGATGAGGAGCAGATG 870
 QY 825 AGCAAAACTCAAGAGAGGAGACCAACCTTATGTATGAGAGGTATCTGCAAGCC 884
 DB 871 AGCAAAACTCAAGAGAGGAGACCAACCTTATGTATGAGAGGTATCTGCAAGCC 930
 QY 885 AGCTGAGATCTTACAGAGCATCCAGTCTTGTCCCAAGGCTGAGAGAGCTACCTGC 944
 DB 931 AGCTGAGATCTTACAGAGCATCCAGTCTTGTCCCAAGGCTGAGAGAGCTACCTGC 990
 QY 945 TTGTGCGCTATGAGAGCTGCTGAGGCCCTGTGAGCCAGACTTCCGAAATGTATGAT 1004

```

Db      |||||
991 TTGTGGCTATAGAGGACCTGGCTCGAGCCCTGTGCCCCAGACTTCCCAATGTATGAAT 1050
OY      1005 TCGTGGATTTGGAATCTTGCCCATCTTCAGACCTGGGTGATATACATACCCGGGGA 1064
Db      1051 TCGTGGATTTGGAATCTTGCCCATCTTCAGACCTGGGTGATATACATACCCGGGGA 1110
OY      1065 AAGGCAATGGGTGACCAAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAG 1124
Db      1111 AAGGCAATGGGTGACCAAGCTTTCACACAAATGCCAGGATGCCCTTAATGTCTCCAG 1170
OY      1125 CTTGGCGCTGTCTTTGGCCCTATGAAAAGTTTCTGACTTCAGAAAAGCTGTGGCGATG 1184
Db      1171 CTTGGCGCTGTCTTTGGCCCTATGAAAAGTTTCTGACTTCAGAAAAGCTGTGGCGATG 1230
OY      1185 CCATGAAATTTGGTGGCTACCCGACAGCTAGATCTGAAACAAGACAGAAACCTGTTC 1244
Db      1231 CCATGAAATTTGGTGGCTACCCGACAGCTAGATCTGAAACAAGACAGAAACCTGTTC 1290
OY      1245 TGGATCTTCTGTCTACCTTGAGCTGTCCGACAAATTCACATAAGGGTTGAGAAGCT 1304
Db      1291 TGGATCTTCTGTCTACCTTGAGCTGTCCGACAAATTCACATAAGGGTTGAGAAGCT 1350
OY      1305 TTGCTGCCACCTGGTGTGACCTGAGCTGCTCTGTAATGCTTGTGAGCCCTGCTAC 1364
Db      1351 TTGCTGCCACCTGGTGTGACCTGAGCTGCTCTGTAATGCTTGTGAGCCCTGCTAC 1410
OY      1365 ATCTCTGACCTTAACTACATGTCTGTGGTATCAGACAGAGTGTGAGTGTGTCACAC 1424
Db      1411 ATCTCTGACCTTAACTACATGTCTGTGGTATCAGACAGAGTGTGAGTGTGTCACAC 1470
OY      1425 GTGCTCAACGACAGAGACTTTGTGTGCAATGCTGTGTCTAGAAAACAGACTGGGAC 1484
Db      1471 GTGCTCAACGACAGAGACTTTGTGTGCAATGCTGTGTCTAGAAAACAGACTGGGAC 1530
OY      1485 TTATGTGACGACACATCCACAGAGAAAGGATTTGCTCTTCTTTCTTTCTGATC 1544
Db      1531 TTATGTGACGACACATCCACAGAGAAAGGATTTGCTCTTCTTTCTTTCTGATC 1590
OY      1545 TTCTGTCTGTGGACACTTCAGAGACTTTGTGGCTGTGAGGCTATTAAGCAGACAG 1604
Db      1591 TTCTGTCTGTGGACACTTCAGAGACTTTGTGGCTGTGAGGCTATTAAGCAGACAG 1650
OY      1605 TATCAGTGAATTTGATCATTAACGTCCTGTCCACATCTTGCCCAATGGGGAATGATC 1664
Db      1651 TATCAGTGAATTTGATCATTAACGTCCTGTCCACATCTTGCCCAATGGGGAATGATC 1710
OY      1665 TTTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1703
Db      1711 TTTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1749

RESULT 5
AAK91803
ID      AAK91803 standard; cDNA; 877 BP.
XX
AC      AAK91803;
XX
DT      06-NOV-2001 (first entry)
XX
DE      Human cDNA 5'-end sequence, SEQ ID NO: 263.
XX
KW      Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
XX
OS      Homo sapiens.
PN      EP130094-A2.
XX
PD      05-SEP-2001.
XX
PF      07-JUL-2000; 2000EP-0114089.
XX
PR      08-JUL-1999; 99JP-0194486.

```

```

PR      11-JAN-2000; 2000JP-0118774.
PR      02-MAY-2000; 2000JP-0183765.
XX
XX      (HELI-) HELIX RES INST.
XX
PI      Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI      Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR      WPI; 2001-524255/58.
XX
PT      830 Primers useful for synthesizing full length cDNA clones and their
PT      use in genetic manipulation.
XX
XX      Claim 2: SEQ ID NO 263; 1380bp + sequence listing; English.
XX
CC      The invention relates to primers for synthesizing full length cDNA
CC      clones. 830 cDNA molecules encoding a human protein have been
CC      isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC      molecules have been determined. Primers for synthesizing the full length
CC      cDNA are useful for clarifying the function of the protein encoded by
CC      the cDNA. The full length clones were obtained by construction of full
CC      length enriched cDNA libraries that were synthesised by the oligo-capping
CC      method. The primers enable the production of the full length cDNA easily
CC      without any special methods. The present sequence is the nucleotide
CC      sequence of the 5'-end of a cDNA provided in the invention.
CC      Note: The sequence data for this patent did not form part of the printed
CC      specification, but was obtained in CD-ROM format directly from EPO.
XX
SO      Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;

Query Match      34.0%; Score 695; DB 22; Length 877;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY      106 CAAGCTCTCCACTTCAGACACAATGCTACTGCTTAATAAATGAAAGCTCTGCTGTTC 165
Db      47 CAAGCTCTCCACTTCAGACACAATGCTACTGCTTAATAAATGAAAGCTCTGCTGTTC 106
OY      166 GGTTCACAGATGGCCATCTTGCTATTTCTTCACATGATACAGCCACATCAGCTC 225
Db      107 GGTTCACAGATGGCCATCTTGCTATTTCTTCACATGATACAGCCACATCAGCTC 166
OY      226 CCGTGTATGAAGGACAGCCGAGGCGATGACAGCGTGTCTCTCTCCGGGGC 285
Db      167 CCGTGTATGAAGGACAGCCGAGGCGATGACAGCGTGTCTCTCTCCGGGGC 226
OY      286 TGGCTCTTTTGTGTGGGAGCTTTTGTGGGAGCAGCCAGATGTTTCTACTGATGGA 345
Db      227 TGGCTCTTTTGTGTGGGAGCTTTTGTGGGAGCAGCCAGATGTTTCTACTGATGGA 286
OY      346 GCCCGCTGGACAGCTGTGATGACTTCAAGCAGACACCGCTGTGATCTGCATGAGC 405
Db      287 GCCCGCTGGACAGCTGTGATGACTTCAAGCAGACACCGCTGTGATCTGCATGAGC 346
OY      406 TGTGGGGGATTCGATACGGGGCGCTCTTGTGGCGCATGAGAGGCTTTGAGGCCAT 465
Db      347 TGTGGGGGATTCGATACGGGGCGCTCTTGTGGCGCATGAGAGGCTTTGAGGCCAT 406
OY      466 GGAACCTGTCTCCCGGAGACAGTCCAGCTCTTTCACTGTGGAGAACAGCCGGCCCTGTG 525
Db      407 GGAACCTGTCTCCCGGAGACAGTCCAGCTCTTTCACTGTGGAGAACAGCCGGCCCTGTG 466
OY      526 TTCTGCACTGCTGTGACATCATCCACACAGATGAATCATCCCGGGGCTCATGACAG 585
Db      467 TTCTGCACTGCTGTGACATCATCCACACAGATGAATCATCCCGGGGCTCATGACAG 526
OY      586 GCTCTGTGCAATGACACACCTTTTGAAGTGTGGAGAGAGGCTTGGCCCTCTACAGCCA 645
Db      527 GCTCTGTGCAATGACACACCTTTTGAAGTGTGGAGAGAGGCTTGGCCCTCTACAGCCA 586
OY      646 CGTGTGCTCAAGAGAGTGCCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAGA 705
Db      587 CGTGTGCTCAAGAGAGTGCCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAGA 646

```

QY	706	CCCCCTCCGCTCAACCTTCATATCGTGCACCTGGTCCGGGACCCCGGGGCGGTTCGGTTC	765
Db	647	CCCCCTCCGCTCAACCTTCATATCGTGCACCTGGTCCGGGACCCCGGGGCGGTTCGGTTC	706
QY	766	CCGAGAACGCACAAAGGAGATCTCATGATTGAC	800
Db	707	CCGAGAACGCACAAAGGAGATCTCATGATTGAC	741

RESULT 6

ID	AAK93921 standard; cDNA; 877 BP
vv	

AC AAK93921;

DT 06-NOV-2001 (first entry)
yy

Human cDNA clone representative sequence, SEQ ID NO: 2381.

KW Human, full length cDNA; cDNA synthesis; oligo-capping; ss.
XY

OS Homo sapiens.

EP1130094-A2

PD 05-SEP-2001.
YY

PF 07-JUL-2000; 2000EP-0114089
YY

PR	08-JUL-1999;	.99JP-0194486.
DP	11-JAN-2000;	3000TB-0118774

PR 02-MAY-2000; 2000JP-0183765.
XX

PA (HELI-) HELIX RES INST.
XX

PI Ota T, Nishikawa T, I
PI Wakamatsu A, Suda Yama

XX
DR WPT: 2001-524255/58

PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
vv

PS Example 11; SEQ ID NO 2381; 1380pp + sequence listing; English.

The invention relates to primers for synthesising full length cDNA clones. 830 cDNA molecules encoding a human protein have been isolated and nucleotide sequences of 5' - and 3' -ends of the cDNA molecules have been determined. Primers for synthesising the full length cDNA are useful for clarifying the function of the protein encoded by the cDNA. The full length clones were obtained by construction of full length enriched cDNA libraries that were synthesised by the oligo-capping method. The primers enable the production of the full length cDNA easily without any special methods. The present sequence was used as the representative sequence from a human clone which was used in the homology searches to identify the clone.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in CD-ROM format directly from EPO.

5Q Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other,

Query Match	34.08;	Score 695;	DB 22;	Length 877;
Post Loca] stml] ex] ttr	100.08;	Docid No. 0;		

Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY	Db	QY	Db
106	47	166	107
CAAGGCTTCCACATTAGACAAATGCTACTGCTTAAAAAATGAAAGCCGTCGGTTTCT	CAAGGCTTCCACATTAGACAAATGCTACTGCTTAAAAAATGAAAGCCGTCGGTTTCT	GGTTTCCCAAGATGCGCATCTTGGCTCTATTCTTCACATGTACAGCCACAAATCAGCTC	GGTTTCCCAAGATGCGCATCTTGGCTCTATTCTTCACATGTACAGCCACAAATCAGCTC
165	106	225	166
CGGCTTCTTCCACATTAGACAAATGCTACTGCTTAAAAAATGAAAGCCGTCGGTTTCT	CGGCTTCTTCCACATTAGACAAATGCTACTGCTTAAAAAATGAAAGCCGTCGGTTTCT	GGTTTCCCAAGATGCGCATCTTGGCTCTATTCTTCACATGTACAGCCACAAATCAGCTC	GGTTTCCCAAGATGCGCATCTTGGCTCTATTCTTCACATGTACAGCCACAAATCAGCTC

OY	226	CCTGCTATGAAAGGACACAGCCGAGGCGCATGACAGCGTGGTTCGTCTTCCTCGGGGGCTC	285
OY	226	CCTGCTATGAAAGGACACAGCCGAGGCGCATGACAGCGTGGTTCGTCTTCCTCGGGGGCTC	285
Db	167	CCTGCTATGAAAGGACACAGCCGAGGCGCATGACAGCGTGGTTCGTCTTCCTCGGGGGCTC	226
OY	286	TGGCTCTCTTTTGTGGGGCAGCTTTTGTGGAGCACCAGATGTTTCTTACCTGATGGA	345
Db	227	TGGCTCTCTTTTGTGGGGCAGCTTTTGTGGAGCACCAGATGTTTCTTACCTGATGGA	286
OY	346	GCCCCCTGGCAGCGTGTGGATGACCTTCAAGCAGAGACACCGCCTGGATCCTGCATYGGC	405
Db	287	GCCCCCTGGCAGCGTGTGGATGACCTTCAAGCAGAGACACCGCCTGGATCCTGCATYGGC	346
OY	406	TGTGGGGGATCGATACGGGGCCGCTCTTGTGGCGCATGAGCGTCTTTGATGCCATCAT	465
Db	347	TGTGGGGGATCGATACGGGGCCGCTCTTGTGGCGCATGAGCGTCTTTGATGCCATCAT	406
OY	466	GGAAACCTGGTCCCGGAGACAGTCCAGCCTCTTTCAGTGGAGAACAGCGGGCCCTGTG	525
Db	407	GGAAACCTGGTCCCGGAGACAGTCCAGCCTCTTTCAGTGGAGAACAGCGGGCCCTGTG	466
OY	526	TTCTGACACCTGGCTGTGACATCATCCACAAAGATGAATCATCCCGGGGCTTCATGACAG	585
Db	467	TTCTGACACCTGGCTGTGACATCATCCACAAAGATGAATCATCCCGGGGCTTCATGACAG	526
OY	586	GCTCTGTGACATCAACAGACCCCTTTGAGGTGGAGGAAGGCGCTGCCGCTCTACAGCCA	645
Db	527	GCTCTGTGACATCAACAGACCCCTTTGAGGTGGAGGAAGGCGCTGCCGCTCTACAGCCA	586
OY	646	CGTGTGCTCAAGGAGAGTGCCTTCTTCAACCTGCAGTCCCTCTACCCGCTGTGAAGA	705
Db	587	CGTGTGCTCAAGGAGAGTGCCTTCTTCAACCTGCAGTCCCTCTTACCCGCTGTGAAGA	646
OY	706	CCCTTCCTCAACCTGACATATGCGACCTGGTCCGGGACCCCGGGGCGTGTCCGTTT	765
Db	647	CCCTTCCTCAACCTGACATATGCGACCTGGTCCGGGACCCCGGGGCGTGTCCGTTT	706
OY	766	CGGAAGACGACAAAGGGAGATCATGATTTGACA	800
Db	707	CGGAAGACGACAAAGGGAGATCATGATTTGACA	741

RESULT 7

ID ABV89280 standard; cDNA; 505 BP.

AC ABV89280

DT 13-DEC-2002 (first entry)

DE Human colon cancer related CDNA SEQ ID NO 2595

KW Human; colon; cancer; cytostatic; tumour; gene therapy; vaccine; gene;

XXXXXX

XX
XX

2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
32
33
34
35
36
37
38
39
40
41
42
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100
101
102
103
104
105
106
107
108
109
110
111
112
113
114
115
116
117
118
119
120
121
122
123
124
125
126
127
128
129
130
131
132
133
134
135
136
137
138
139
140
141
142
143
144
145
146
147
148
149
150
151
152
153
154
155
156
157
158
159
160
161
162
163
164
165
166
167
168
169
170
171
172
173
174
175
176
177
178
179
180
181
182
183
184
185
186
187
188
189
190
191
192
193
194
195
196
197
198
199
200
201
202
203
204
205
206
207
208
209
210
211
212
213
214
215
216
217
218
219
220
221
222
223
224
225
226
227
228
229
230
231
232
233
234
235
236
237
238
239
240
241
242
243
244
245
246
247
248
249
250
251
252
253
254
255
256
257
258
259
260
261
262
263
264
265
266
267
268
269
270
271
272
273
274
275
276
277
278
279
280
281
282
283
284
285
286
287
288
289
290
291
292
293
294
295
296
297
298
299
300
301
302
303
304
305
306
307
308
309
310
311
312
313
314
315
316
317
318
319
320
321
322
323
324
325
326
327
328
329
330
331
332
333
334
335
336
337
338
339
340
341
342
343
344
345
346
347
348
349
350
351
352
353
354
355
356
357
358
359
360
361
362
363
364
365
366
367
368
369
370
371
372
373
374
375
376
377
378
379
380
381
382
383
384
385
386
387
388
389
390
391
392
393
394
395
396
397
398
399
400
401
402
403
404
405
406
407
408
409
410
411
412
413
414
415
416
417
418
419
420
421
422
423
424
425
426
427
428
429
430
431
432
433
434
435
436
437
438
439
440
441
442
443
444
445
446
447
448
449
450
451
452
453
454
455
456
457
458
459
460
461
462
463
464
465
466
467
468
469
470
471
472
473
474
475
476
477
478
479
480
481
482
483
484
485
486
487
488
489
490
491
492
493
494
495
496
497
498
499
500
501
502
503
504
505
506
507
508
509
510
511
512
513
514
515
516
517
518
519
520
521
522
523
524
525
526
527
528
529
530
531
532
533
534
535
536
537
538
539
540
541
542
543
544
545
546
547
548
549
550
551
552
553
554
555
556
557
558
559
560
561
562
563
564
565
566
567
568
569
570
571
572
573
574
575
576
577
578
579
580
581
582
583
584
585
586
587
588
589
590
591
592
593
594
595
596
597
598
599
600
601
602
603
604
605
606
607
608
609
610
611
612
613
614
615
616
617
618
619
620
621
622
623
624
625
626
627
628
629
630
631
632
633
634
635
636
637
638
639
640
641
642
643
644
645
646
647
648
649
650
651
652
653
654
655
656
657
658
659
660
661
662
663
664
665
666
667
668
669
670
671
672
673
674
675
676
677
678
679
680
681
682
683
684
685
686
687
688
689
690
691
692
693
694
695
696
697
698
699
700
701
702
703
704
705
706
707
708
709
710
711
712
713
714
715
716
717
718
719
720
721
722
723
724
725
726
727
728
729
730
731
732
733
734
735
736
737
738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763
764
765
766
767
768
769
770
771
772
773
774
775
776
777
778
779
780
781
782
783
784
785
786
787
788
789
790
791
792
793
794
795
796
797
798
799
800
801
802
803
804
805
806
807
808
809
810
811
812
813
814
815
816
817
818
819
820
821
822
823
824
825
826
827
828
829
830
831
832
833
834
835
836
837
838
839
840
841

XX 10-NOV-2003
DE

XX 20-NOV-2000 200015-252222P

PR 06-FEB-2001; 2001US-267011P
PR 08-MAR-2001; 2001US-270670P

PR 10-JUL-2001; 2001US-304037P
YY

PA (CORI-) CORIXA CORP.
YY

PI Stolck JA, Xu J, Chenault H
YV

DR WPI; 2002-608400/65

XX New isolated tumor colon polynucleotide and polypeptide, useful for the
 PT diagnosis, prevention and/or treatment of cancer, in particular colon
 PT cancer

PS Claim 1: SEQ ID NO 2595; 266pp + Sequence Listing; English.

XX The invention relates to a human colon tumour expressed polynucleotide
 CC (I) encoding a polypeptide (II), ABP67991-ABP67996) comprising: (i) any of
 CC 2600 fully defined nucleotide sequences (ABV8669-ABV89289); (ii)
 CC complements of (i); (iii) at least 20 contiguous residues of (i); (iv)
 CC sequences that hybridize to (i), under moderately stringent conditions;
 CC (v) sequences having at least 75% or 90% identity to (i); or (vi)
 CC degenerate variants of (i). The compositions and methods of the present
 CC invention are useful for the diagnosis, prevention and/or treatment of
 CC cancer, particularly colon cancer. (i) can be used in gene therapy and
 CC (I) and (II) are useful in pharmaceutical compositions such as vaccines.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at ftp.wipo.int/pub/published_pcl_sequences.

XX Sequence 505 BP; 120 A; 125 C; 119 G; 141 T; 0 other;

XX Query Match 19.7%; Score 403; DB 24; Length 505;
 XX Best Local Similarity 100.0%; Pred. No. 2.6e-189;

XX Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCCTTGGCTGCGACCTGCTGAGCCTTCTGTAATGCTTCTGAGCCTTGC 1360

DB 1 GGCCTTGGCTGCGACCTGCTGAGCCTTCTGTAATGCTTCTGAGCCTTGC 60

QY 1361 CTACATCTCTGAGCCTTAACTACATGCTGTGGGATACACAGTGTGAGTGTGTC 1420

DB 61 CTACATCTCTGAGCCTTAACTACATGCTGTGGGATACACAGTGTGAGTGTGTC 120

QY 1421 ACAGTGTCTCAAGCAGAAAGAGCTTGTGTCATGCTTGTCTGTAATGCTTCTGAGCCTTGC 1480

DB 121 ACAGTGTCTCAAGCAGAAAGAGCTTGTGTCATGCTTGTCTGTAATGCTTCTGAGCCTTGC 180

QY 1481 AACCTTATGTGAGCAGACATCCACAGTGAAGAGGATATGCTTCTTCTTCTTCTT 1540

DB 181 AACCTTATGTGAGCAGACATCCACAGTGAAGAGGATATGCTTCTTCTTCTTCTTCTT 240

QY 1541 GATCTTCTGCTGGGAGACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1600

DB 241 GATCTTCTGCTGGGAGACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 300

QY 1601 ACAGTATCAGTGAATGATCCATAAACCCTGCTCCACATCTTGCCCAATGGGGAATG 1660

DB 301 ACAGTATCAGTGAATGATCCATAAACCCTGCTCCACATCTTGCCCAATGGGGAATG 360

QY 1661 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1703

DB 361 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 403

QY 1661 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1703

DB 361 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 403

QY 1661 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1703

DB 361 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 403

QY 1661 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1703

DB 361 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 403

QY 1661 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1703

DB 361 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 403

QY 1661 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1703

DB 361 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 403

QY 1661 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1703

DB 361 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 403

QY 1661 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1703

DB 361 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 403

QY 1661 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1703

DB 361 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 403

XX 30-JUL-2001; 2001WO-US23826.

XX 03-AUG-2000; 2000US-223265P.

XX 02-OCT-2000; 2000US-237406P.

XX 20-MAR-2001; 2001US-277495P.

XX 03-JUL-2001; 2001US-302702P.

XX (CORI-) CORIXA CORP.

XX Pyle RA, Xu J, Secretist H;

XX WPI; 2002-257462/30.

XX Novel polynucleotide encoding colon tumour polypeptides, useful as

XX vaccines for treating colon cancers

XX Claim 1: Page 206; 425pp; English.

XX The invention relates to isolated polynucleotides (I) encoding colon

XX tumour polypeptides (II). (I) is useful for stimulating an immune

XX response in a patient and treating colon cancer in a patient.

XX Oligonucleotides derived from (I) are useful for determining the presence

XX of cancer in a patient. (I) and (II) are useful in pharmaceutical

XX compositions, e.g. vaccines, and other compositions for the diagnosis

XX and treatment of colon cancer. A composition comprising a first component

XX selected from physiologically acceptable carriers and immunostimulants,

XX and an antigen-presenting cell expressing (II) is useful for inhibiting

XX development of cancer in a patient. (I) is useful in the design and

XX preparation of ribozyme molecules for inhibiting expression of tumour

XX polypeptides and (II). AAK5431-ABK55464 represent human colon cancer cDNA

XX sequences of the invention.

XX Sequence 517 BP; 121 A; 130 C; 124 G; 142 T; 0 other;

XX Query Match 19.7%; Score 403; DB 24; Length 517;
 XX Best Local Similarity 100.0%; Pred. No. 2.6e-189;

XX Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCCTTGGCTGCGACCTGCTGAGCCTTCTGTAATGCTTCTGAGCCTTGC 1360

DB 13 GGCCTTGGCTGCGACCTGCTGAGCCTTCTGTAATGCTTCTGAGCCTTGC 72

QY 1361 CTACATCTCTGAGCCTTAACTACATGCTGTGGGATACACAGTGTGAGTGTGTC 1420

DB 73 CTACATCTCTGAGCCTTAACTACATGCTGTGGGATACACAGTGTGAGTGTGTC 132

QY 1421 ACAGTGTCTCAAGCAGAAAGAGCTTGTGTCATGCTTGTCTGTAATGCTTCTGAGCCTTGC 1480

DB 133 ACAGTGTCTCAAGCAGAAAGAGCTTGTGTCATGCTTGTCTGTAATGCTTCTGAGCCTTGC 192

QY 1481 AACCTTATGTGAGCAGACATCCACAGTGAAGAGGATATGCTTCTTCTTCTTCTT 1540

DB 199 AACCTTATGTGAGCAGACATCCACAGTGAAGAGGATATGCTTCTTCTTCTTCTTCTT 252

QY 1541 GATCTTCTGCTGGGAGACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 1600

DB 253 GATCTTCTGCTGGGAGACTTTCAGAGACTTGTGGCTGAGGCTTATTAAGCAGAC 312

XX 1601 ACAGTATCAGTGAATGATCCATAAACCCTGCTCCACATCTTGCCCAATGGGGAATG 1660

XX 313 ACAGTATCAGTGAATGATCCATAAACCCTGCTCCACATCTTGCCCAATGGGGAATG 372

XX 1661 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 1703

XX 373 GATCTTTCACCAAGAGCTACACAGCATTTTCCACAGAGATGC 415

XX RESULT 9

XX AAC76156 standard; cDNA; 2988 BP.

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX AAC76156

XX 08-FEB-2001 (first entry)
 DT Human ORFX ORF1711 polynucleotide sequence SEQ ID NO:3421.
 XX
 DE
 XX Human: open reading frame: ORFX: detection; cytostatic; hepatotropic;
 KM vlnenary; antiposrotic; antiparkinsonian; nootropic; neuroprotective;
 KM anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
 KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KM hypotensive; dermatological; immunosuppressive; antineoplastic;
 KM antitumor; antibacterial; antifungal; antirheumatic; antihypertensive;
 KM antidiabetic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KM thrombosis; contraceptive; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200058473-A2.
 XX
 D 05-OCT-2000.
 XX
 PF 31-MAR-2000; 2000MO-US08621.
 XX
 PR 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 XX (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 XX
 DR WPI; 2000-602362/57.
 DR P-PSDB; AAB41947.
 XX
 XX Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 5; Page 2597-2599; 5507Pp; English.
 XX
 CC AAC74446 to AAC7606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vlnenary;
 CC antiposrotic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiarthritic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antihypertensive; antidiabetic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy.
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 XX
 S0 Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other;
 Query Match 15.0%; Score 306; DB 21; Length 2988;
 Best Local Similarity 100.0%; Pred. No. 3,7e-141;
 Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 452 TTGTATGCTACATGAGACCTGGTCCCGAGAGACGTCAGCCCTTTTCAGTGAGAGAC 511
 DB 1480 TTGTATGCTACATGAGACCTGGTCCCGAGAGACGTCAGCCCTTTTCAGTGAGAGAC 1539
 QY 512 AGCCGGGCGCTGTGTTGTCGACCTGCTGTGACATATCCACAGATGAATCATCC 571
 DB 1540 AGCCGGGCGCTGTGTTGTCGACCTGCTGTGACATATCCACAGATGAATCATCC 1599
 QY 572 CGGGCTACATGACGCTCCCTGTGTCAGTCACAGCCCTTTAGGTGGGGAAGGCGTGC 631
 DB 1600 CGGGCTACATGACGCTCCCTGTGTCAGTCACAGCCCTTTAGGTGGGGAAGGCGTGC 1659
 QY 632 CGGTCTACAGCCAGCGTGTGTCAGAGAGAGTGGGCTTTCAACCTGACGTCCTTAC 691
 DB 1660 CGGTCTACAGCCAGCGTGTGTCAGAGAGAGTGGGCTTTCAACCTGACGTCCTTAC 1719
 QY 692 CGGTCTGTAAGACCCCTCCCTCAACCTGATATGTCGACCTGTCGGGAGACCCCGG 751
 DB 1720 CGGTCTGTAAGACCCCTCCCTCAACCTGATATGTCGACCTGTCGGGAGACCCCGG 1779
 QY 752 GCGGTG 757
 DB 1780 GCGGTG 1785
 RESULT 10
 ABR54794
 ID ABR54794 standard; cDNA; 389 BP.
 XX
 AC ABR54794;
 XX
 DT 18-JUN-2002 (first entry)
 XX
 DE Human colon cancer-associated cDNA, SEQ ID NO 264.
 XX
 KM Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.
 XX
 OS Homo sapiens.
 XX
 PN WO200212280-A2.
 XX
 PD 14-FEB-2002.
 XX
 PF 30-JUL-2001; 2001MO-US23826.
 XX
 PR 03-AUG-2000; 2000US-23265P.
 PR 02-OCT-2000; 2000US-237406P.
 PR 20-MAR-2001; 2001US-277495P.
 PR 03-JUL-2001; 2001US-302702P.
 XX
 PA (CORI-) CORIXA CORP.
 XX
 PI Pyle RA, Xu J, Secrist H;
 XX
 DR WPI; 2002-257462/30.
 XX
 PT Novel polynucleotide encoding colon tumour polypeptides, useful as
 PT vaccines for treating colon cancers -
 XX
 PS Claim 1; Page 225; 425Pp; English.
 XX
 XX The invention relates to isolated polynucleotides (I) encoding colon
 CC tumour polypeptides (II). (I) is useful for stimulating an immune
 CC response in a patient and treating colon cancer in a patient.
 CC Oligonucleotides derived from (I) are useful for determining the presence
 CC of cancer in a patient. (I) and (II) are useful in pharmaceutical
 CC compositions, e.g. vaccines, and other compositions for the diagnosis
 CC and treatment of colon cancer. A composition comprising a first component
 CC selected from physiologically acceptable carriers and immunostimulants,
 CC and an antigen-presenting cell expressing (II) is useful for inhibiting
 CC development of cancer in a patient. (I) is useful in the design and
 CC preparation of ribozyme molecules for inhibiting expression of tumour

CC polypeptides and (I). ABR54531-ABR55464 represent human colon cancer cDNA
sequences of the invention.

XX Sequence 389 BP; 121 A; 83 C; 76 G; 109 T; 0 other;

Query Match 5.6%; Score 114; DB 24; Length 389;
Best Local Similarity 100.0%; Pred. No. 6.9e-46;
Matches 114; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1930 TGCCATCTGTAACTAAATCCCAATAGGTCGTTAGATGTCCTTTTATG 1989

DB 266 TGCCATCTGTAACTAAATCCCAATAGGTCGTTAGATGTCCTTTTATG 325

OY 1990 CTTCTTAATATTAGCAGTAATGTTTCATTTTATGGATCCTAATAAAAAA 2043
DB 326 CTTCTTAATATTAGCAGTAATGTTTCATTTTATGGATCCTAATAAAAAA 379

RESULT 11

AAK92588/c
ID AAK92588 standard; cDNA; 548 BP.

AC AAK92588;

DE 06-NOV-2001 (first entry)

DE Human cDNA 3'-end sequence, SEQ ID NO: 1048.

KW Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

OS Homo sapiens.

PN EP1130094-A2.

PD 05-SEP-2001.

PF 07-JUL-2000; 2000EP-0114089.

PR 08-JUL-1999; 99JP-0194486.

PR 11-JAN-2000; 2000JP-0118774.

PR 02-MAY-2000; 2000JP-0183765.

XX (HELI-) HELIX RES INST.

XX Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y,
XX Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI; 2001-524255/58.

DR 830 Primers useful for synthesizing full length cDNA clones and their
XX use in genetic manipulation -

XX Claim 3; SEQ ID NO 1048; 1380pp + sequence listing; English.

CC The invention relates to primers for synthesizing full length cDNA
XX clones. 830 cDNA molecules encoding a human protein have been

CC isolated and nucleotide sequences of 5' and 3' ends of the cDNA
XX molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by

CC the cDNA. The full length clones were obtained by construction of full
XX length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily

CC without any special methods. The present sequence is the nucleotide
XX sequence of the 3'-end of a cDNA provided in the invention.

CC Note: The sequence data for this patent did not form part of the printed
XX specification, but was obtained in CD-ROM format directly from EPO.

CC Sequence 548 BP; 160 A; 109 C; 129 G; 148 T; 2 other;

XX Query Match 5.1%; Score 105; DB 22; Length 548;

XX Best Local Similarity 100.0%; Pred. No. 2e-41;

XX Matches 105; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1930 TGCCATCTGTAACTAAATCCCAATAGGTCGTTAGATGTCCTTTTATG 1989

DB 108 TGCCATCTGTAACTAAATCCCAATAGGTCGTTAGATGTCCTTTTATG 49

OY 1990 CTTCTTAATATTAGCAGTAATGTTTCATTTTATGGATCCTAA 2034

DB 48 CTTCTTAATATTAGCAGTAATGTTTCATTTTATGGATCCTAA 4

RESULT 12

AAAD24670
ID AAD24670 standard; cDNA; 1647 BP.

AC AAD24670;

DE 12-MAR-2002 (first entry)

DE Human drug metabolizing enzyme (DME)-5 cDNA.

KW Human; drug metabolizing enzyme; gene therapy; autoimmune disorder;
KW inflammatory disorder; acquired immune deficiency syndrome; infection;
KW AIDS; adult respiratory distress syndrome; cell proliferative disorder;
KW allergy; anaemia; conjunctivitis; actinic keratosis; arteriosclerosis;
KW cancer; endocrine disorder; hypothalamus disorder; pituitary disorder;
KW gastrointestinal disorder; metabolic disorder; developmental disorder;
KW liver disorder; iritis; cystic fibrosis; Addison's disease; retinitis;
KW golfer; diabetes; dysphagia; gastric carcinoma; anorexia; eye disorder;
KW DME-5; ss.

OS Homo sapiens.

PN Key Location/Qualifiers

FT CDS 1..1188

FT sig_peptide /tag= a

FT mat_peptide /tag= b

FT sig_peptide /tag= c

FT sig_peptide /product= "Human mature DME-5 protein #1"

FT mat_peptide /tag= d

FT sig_peptide /tag= e

FT mat_peptide /product= "Human mature DME-5 protein #2"

FT WO200179468-A2.

PN 25-OCT-2001.

PD 12-APR-2001; 2001WO-US11869.

PF 13-APR-2000; 2000US-197590P.

PR 19-APR-2000; 2000US-198403P.

PR 28-APR-2000; 2000US-200185P.

PR 05-MAY-2000; 2000US-202234P.

PR 11-MAY-2000; 2000US-203509P.

XX (INCY-) INCYTE GENOMICS INC.

XX Policky JL, Hafalia A, Burford N, Ring HZ, Lal P, Tribouley CM,
XX Yao MG, Yue H, Tang YT, Patterson C, Das D, Sanjanwala MS,
XX Gandhi AR, Reddy R, Khan FA, Baughn MR, Ramkumar J, Griffin JA;
XX Au-Yang J;

XX WPI: 2002-066363/09.

XX P-PSDB: AAE15438.

XX Novel isolated human drug metabolizing enzymes referred as DME 1-10,
XX useful for diagnosing, treating, or preventing disorders associated
XX with aberrant expression of DME such as allergy, anemia, asthma,

XX infertility -

PS Claim 5; Page 139; 143pp; English.

XX The invention relates to human drug metabolizing enzymes referred as
CC DME and nucleic acid molecules encoding such enzymes. Polynucleotides
CC of the invention are useful for assessing toxicity of test compounds
CC and in gene therapy. Sequences of the invention are useful in the
CC diagnosis, prevention and treatment of autoimmune/inflammatory
CC disorders such as acquired immune deficiency syndrome (AIDS), adult
CC respiratory distress syndrome, allergies, anaemia, atherosclerosis,
CC asthma, autoimmune haemolytic anaemia, contact dermatitis, Crohn's
CC disease, glomerulonephritis, Goodpasture's syndrome, gout, Grave's
CC disease, Hashimoto's thyroiditis, irritable bowel syndrome, multiple
CC sclerosis, osteoarthritis, osteoporosis, psoriasis, systemic lupus
CC erythematosus, rheumatoid arthritis, scleroderma, ulcerative colitis,
CC uveitis, viral, bacterial, fungal, parasitic, protozoal, helminthic
CC infections; cell proliferative disorders such as actinic keratosis,
CC arteriosclerosis, atherosclerosis, Duchenne and Becker dystrophy,
CC hepatitis, Cushing's syndrome, cancers, myelodysplastic syndrome,
CC epilepsy; endocrine disorders such as disorders of the hypothalamus
CC and pituitary resulting from lesions such as primary brain tumours,
CC adenomas, infarction associated with pregnancy, aneurysms, vascular
CC malformations; eye disorders such as conjunctivitis, iritis, retinitis,
CC glaucoma, pigmentosa, metabolic disorders such as Addison's disease,
CC cystic fibrosis, diabetes, goitre, glycogen storage diseases,
CC hypercholesterolaemia, hyperthyroidism, hypoglycaemia, lipid myopathies,
CC Menkes syndrome, mannosidosis, obesity; gastrointestinal disorders such
CC as dysphagia, gastric carcinoma, anorexia, nausea, gastroenteritis,
CC hyperbilirubinaemia, emesis, cirrhosis, diarrhoea, jaundice, Reye's
CC syndrome, peliosis hepatitis, hepatic vein thrombosis and developmental
CC disorders. The present sequence is human DME-5 protein cDNA.

XX
SQ Sequence 1647 BP; 287 A; 560 C; 501 G; 299 T; 0 other;

Query Match 2.9%; Score 59; DB 24; Length 1647;
Best Local Similarity 100.0%; Pred. No. 1.4e-18;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 624 AGGCTGCCGCTCTACAGCCAGCGTGTCTCAAGAGAGTGCGCTTCTCAACCTGCAG 682
|||||
DB 488 AGGCTGCCGCTCTACAGCCAGCGTGTCTCAAGAGAGTGCGCTTCTCAACCTGCAG 546

RESULT 13
AAD02700
ID AAD02700 standard; cDNA; 1694 BP.

AC AAD02700;
XX
DT 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-4beta (GST-4beta) cDNA.

DE Human: glycosyl sulfotransferase-4beta; GST-4beta; Immunosuppressive;
XX therapy; selectin binding inhibitor; gene therapy; inflammation;
XX systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
XX polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
XX glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;
XX Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
XX demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
XX myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
XX asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX chromosome 16q23.1; ss.

OS Homo sapiens.
XX
FH Key Location/Qualifiers
FT 5'UTR 9..188
FT CDS 189..1376
FT /tag- a
FT /tag- b
FT /product- "Human glycosyl sulfotransferase-4beta
FT (GST-4beta)"
FT /note- "CDS is referred as SEQ ID NO:21 in brief

FT description of the figures (page no.: 4)"
FT 3'UTR 1377..1694
FT /**tag- c
XX
XX WO200106015-A1.
XX
XX 25-JAN-2001.
XX
XX 19-JUL-2000; 2000WO-US19741.
XX
XX 20-JUL-1999; 99US-0144694.
XX
XX 13-JUL-2000; 2000US-0593828.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Rosen SD, Lee JK, Hemmerich S;
XX WPI; 2001-138471/14.
XX P-PSDB; NAY72640.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications
XX
XX Claim 6; Fig 4A; 128pp; English.

XX The present sequence is human glycosyl sulfotransferase-4beta (GST-4
XX beta) cDNA. GST-4 gene is found on chromosome 16q23.1.
XX GST is a type 2 membrane protein useful for inhibiting a binding event
XX between a selectin and a selectin ligand, which comprises contacting the
XX agent that inhibits the sulphation activity of GST. GST is also useful
XX in inhibiting a selectin mediated binding event. GST is useful in gene
XX therapy to treat disorders such as acute or chronic inflammation,
XX systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
XX nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX disease, Grave's disease, adrenailitis, hypoparathyroidism, pernicious
XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX during transplantation.

XX
SQ Sequence 1694 BP; 259 A; 586 C; 514 G; 335 T; 0 other;

Query Match 2.9%; Score 59; DB 22; Length 1694;
Best Local Similarity 100.0%; Pred. No. 1.4e-18;
Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 624 AGGCTGCCGCTCTACAGCCAGCGTGTCTCAAGAGAGTGCGCTTCTCAACCTGCAG 682
|||||
DB 676 AGGCTGCCGCTCTACAGCCAGCGTGTCTCAAGAGAGTGCGCTTCTCAACCTGCAG 734

RESULT 14
AAD02699
ID AAD02699 standard; cDNA; 2044 BP.

AC AAD02699;
XX
DT 02-MAY-2001 (first entry)

XX Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #2.

DE Human: glycosyl sulfotransferase-4alpha; GST-4alpha; Immunosuppressive;
XX therapy; selectin binding inhibitor; gene therapy; inflammation;
XX systemic lupus erythematosus; SLE; rheumatoid arthritis; diabetes;
XX polyarteritis nodosa; polymyositis; systemic sclerosis; dermatitis;
XX glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;
XX Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
XX demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
XX myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
XX asthma; hypersensitivity; rheumatic fever; tissue rejection;

KW chromosome 16q23.1; ss.
XX
XX Homo sapiens.
XX
XX Key Location/Qualifiers
XX CDS 218..1390
XX /tag= a
XX /product= "Human glycosyl sulfotransferase-4alpha
XX (GST-4alpha)"
XX /note= "CDS is specifically claimed as SEQ ID NO: 4
XX in claim 6 (page no: 41) of the specification"
XX
XX WO200106015-A1.
XX
XX PD 25-JAN-2001.
XX
XX 19-JUL-2000; 2000MO-US19741.
XX
XX 20-JUL-1999; 99US-0144694.
XX 13-JUL-2000; 2000US-0593828.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Rosen SD, Lee JK, Hemmerich S;
XX
XX WPI: 2001-138471/14.
XX P-PSDB; AAY72639.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications
XX
XX Claim 6; Fig 1; 128pp; English.
XX
XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
XX alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.
XX GST is a type 2 membrane protein useful for inhibiting a binding event
XX between a selectin and a selectin ligand, which comprises contacting the
XX selectin with a non-sulphated selectin ligand, GST and a small molecular
XX agent that inhibits the sulphation activity of GST. GST is also useful
XX in inhibiting a selectin mediated binding event. GST is useful in gene
XX therapy to treat disorders such as acute or chronic inflammation,
XX systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis
XX nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX disease, Grave's disease, adrenailitis, hypoparathyroidism, pernicious
XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX during transplantation.
XX Note: The present sequence is also shown in sequence listing (page
XX no: 56) but lacks four nucleotides at its 3' end.
XX
XX Sequence 2044 BP; 422 A; 645 C; 570 G; 407 T; 0 other;
XX
XX Query Match 2.9%; Score 59; DB 22; Length 2044;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-18;
XX Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 624 AGGCTGCGGCTCTACAGCCAGTGCTCAAGAGAGTGCGCTTCAACCTGCAG 682
XX |||||||
XX DB 708 AGGCTGCGGCTCTACAGCCAGTGCTCTCAAGAGAGTGCGCTTCAACCTGCAG 766
XX |||||||
XX
XX RESULT 15
XX AAD02698
XX ID AAD02698 standard; cDNA: 2170 BP.
XX
XX AC AAD02698;
XX
XX DT 02-MAY-2001 (first entry)
XX
XX DE Human glycosyl sulfotransferase-4alpha (GST-4alpha) cDNA #1.

XX
XX Human; glycosyl sulfotransferase-4alpha; GST-4alpha; immunosuppressive;
XX therapy; selectin binding inhibitor; gene therapy; inflammation;
XX systemic lupus erythematosus, SLE, rheumatoid arthritis, diabetes;
XX polyarthritis nodosa; polymyositis; systemic sclerosis; dermatitis;
XX glomerulonephritis; myasthenia gravis; Sjogren's syndrome; adrenailitis;
XX Hashimoto's disease; Grave's disease; hypoparathyroidism; anaemia;
XX demyelinating disease; cirrhosis; ulcerative colitis; allergic rhinitis;
XX myocarditis; adult respiratory distress syndrome; eczema; psoriasis;
XX asthma; hypersensitivity; rheumatic fever; tissue rejection;
XX chromosome 16q23.1; ss.
XX
XX OS Homo sapiens.
XX
XX XX
XX Key Location/Qualifiers
XX 5'UTR 9..343
XX /tag= a
XX CDS 344..1516
XX /tag= b
XX /product= "Human glycosyl sulfotransferase-4alpha
XX (GSR-4alpha)"
XX /note= "CDS is specifically claimed as SEQ ID NO: 4
XX in claim 6 (page no: 41) of the specification"
XX
XX 3'UTR 1517..2134
XX /tag= c
XX
XX WO200106015-A1.
XX
XX PD 25-JAN-2001.
XX
XX 19-JUL-2000; 2000MO-US19741.
XX
XX 20-JUL-1999; 99US-0144694.
XX 13-JUL-2000; 2000US-0593828.
XX
XX (REGC) UNIV CALIFORNIA.
XX
XX Rosen SD, Lee JK, Hemmerich S;
XX
XX WPI: 2001-138471/14.
XX P-PSDB; AAY72639.
XX
XX New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX diagnostic and therapeutic agent screening applications
XX
XX Claim 6; Page 62; 128pp; English.
XX
XX The present sequence is human glycosyl sulfotransferase-4alpha (GST-4
XX alpha) cDNA. GST-4 gene is found on chromosome 16q23.1.
XX GST is a type 2 membrane protein useful for inhibiting a binding event
XX between a selectin and a selectin ligand, which comprises contacting the
XX selectin with a non-sulphated selectin ligand, GST and a small molecular
XX agent that inhibits the sulphation activity of GST. GST is also useful
XX in inhibiting a selectin mediated binding event. GST is useful in gene
XX therapy to treat disorders such as acute or chronic inflammation,
XX systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarthritis
XX nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
XX glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
XX disease, Grave's disease, adrenailitis, hypoparathyroidism, pernicious
XX anaemia, demyelinating diseases, cirrhosis, ulcerative colitis,
XX dermatitis, myocarditis, regional enteritis, adult respiratory distress
XX syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
XX bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
XX during transplantation.
XX
XX Sequence 2170 BP; 449 A; 682 C; 612 G; 427 T; 0 other;
XX
XX Query Match 2.9%; Score 59; DB 22; Length 2170;
XX Best Local Similarity 100.0%; Pred. No. 1.4e-18;
XX Matches 59; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 624 AGGCTGCGGCTCTACAGCCAGTGCTCAAGAGAGTGCGCTTCAACCTGCAG 682
XX |||||||

Fri Aug 15 08:09:33 2003

us-09-645-078-1_3.rng

Page 13

Db 834 AGGCTGCGCTCCTACAGCCACGTGGTCTCAAGAGAGTGCCTTCTTCANACCTGCAG 892

Search completed: August 14, 2003, 18:09:57
Job time : 544 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 15:34:51 ; Search time 7280 Seconds

(without alignments)
11480.545 Million cell updates/sec

Title: US-09-645-078-1

Perfect score: 2043

Sequence: 1 gaattcattgtgttggtta.....tgggatacctaaataaaaaa 2043

Scoring table: OLIGO_NUC

Gapop 60.0 , Gapext 60.0

Word size : 0

Total number of hits satisfying chosen parameters: 5770842

Minimum DB seq length: 10

Maximum DB seq length: 2000000000

Post-processing: Listing first 45 summaries

Database :

GenBank:*

1: gb-ba:*

2: gb-ba:*

3: gb-ba:*

4: gb-ba:*

5: gb-ba:*

6: gb-ba:*

7: gb-ba:*

8: gb-ba:*

9: gb-ba:*

10: gb-ba:*

11: gb-ba:*

12: gb-ba:*

13: gb-ba:*

14: gb-ba:*

15: gb-ba:*

16: gb-ba:*

17: gb-ba:*

18: gb-ba:*

19: gb-ba:*

20: gb-ba:*

21: gb-ba:*

22: gb-ba:*

23: gb-ba:*

24: gb-ba:*

25: gb-ba:*

26: gb-ba:*

27: gb-ba:*

28: gb-ba:*

29: gb-ba:*

30: gb-ba:*

31: gb-ba:*

32: gb-ba:*

33: gb-ba:*

34: gb-ba:*

35: gb-ba:*

36: gb-ba:*

37: gb-ba:*

38: gb-ba:*

39: gb-ba:*

40: gb-ba:*

41: gb-ba:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1722	84.3	1992	9	AF280088 Homo sapi
2	1653	80.9	2037	9	BC035282 Homo sapi
3	1598	78.2	2011	9	AK026635 Homo sapi
4	1597	78.2	183228	9	AC010547 Homo sapi
5	1597	78.2	206943	9	AC138848 Homo sapi
6	1547	75.7	1979	6	BD127258 Homo sapi
7	1547	75.7	1979	6	AK074746 Homo sapi
8	1547	75.7	2032	6	AR033335 Sequence
9	1547	75.7	2032	6	BD134772 Glycosyls
10	1547	75.7	2032	6	AF131235 Homo sapi
11	1547	75.7	2032	6	AF149783 Homo sapi
12	1547	75.7	2032	6	AR310485 Sequence
13	1547	75.7	2032	6	BD124832 primer fo
14	1547	75.7	2032	6	BD126950 primer fo
15	1547	75.7	2032	6	AX381256 Sequence
16	1547	75.7	2032	6	AX381326 Sequence
17	1547	75.7	2032	6	BD125617 primer fo
18	1547	75.7	2032	6	AF176839 Homo sapi
19	1547	75.7	2032	6	AX327330 Sequence
20	1547	75.7	2032	6	AF176838 Homo sapi
21	1547	75.7	2032	6	AF219990 Homo sapi
22	1547	75.7	2032	6	AF246718 Homo sapi
23	1547	75.7	2032	6	AF280086 Homo sapi
24	1547	75.7	2032	6	AF219991 Homo sapi
25	1547	75.7	2032	6	AC009163 Homo sapi
26	1547	75.7	2032	6	AC025287 Homo sapi
27	1547	75.7	2032	6	AC009105 Homo sapi
28	1547	75.7	2032	6	AR203340 Sequence
29	1547	75.7	2032	6	BD134774 Glycosyls
30	1547	75.7	2032	6	BD134553 Method fo
31	1547	75.7	2032	6	BD134773 Glycosyls
32	1547	75.7	2032	6	AF131236 Mus muscu
33	1547	75.7	2032	6	AF109155 Mus muscu
34	1547	75.7	2032	6	AC130174 Rattus no
35	1547	75.7	2032	6	AC110314 Rattus no
36	1547	75.7	2032	6	AC121740 Rattus no
37	1547	75.7	2032	6	AC133449 Rattus no
38	1547	75.7	2032	6	AC096328 Rattus no
39	1547	75.7	2032	6	AC095651 Rattus no
40	1547	75.7	2032	6	AC137365 Rattus no
41	1547	75.7	2032	6	AC129764 Rattus no
42	1547	75.7	2032	6	AC094919 Rattus no
43	1547	75.7	2032	6	AF176841 Mus muscu
44	1547	75.7	2032	6	AF176840 Mus muscu
45	1547	75.7	2032	6	AC099262_1 Continuation (2 of

ALIGNMENTS

RESULT 1

AF280088

LOCUS

DEFINITION Homo sapiens l-selectin ligand sulfotransferase GST-3 mRNA, complete cds.

ACCESSION AF280088

VERSION AF280088.1

KEYWORDS

SOURCE

ORGANISM Homo sapiens (human)

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1992)

Hemmerich,S., Lee,J.K., Bhakta,S., Bistrup,A., Ruddle,N.R. and Rosen,S.D.

Pred. No. is the number of results predicted by chance to have a

TITLE Chromosomal localization and genomic organization for the galactose/N-acetylglucosamine/N-acetylglucosamine 6-O-sulfotransferase gene family

JOURNAL Glycobiology 11 (1), 75-87 (2001)

MEDLINE 21096027

PubMed 11181564

REFERENCE 2 (bases 1 to 1992)

AUTHORS Hemmerlich, S., Bhakta, S., Lee, J.-K., Bistrup, A., Ruddie, N.R. and Rosen, S.D.

TITLE Direct Submission

JOURNAL Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience, 3401 Hillview Ave., Palo Alto, CA 94304, USA

FEATURES

source

1. 1992

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/chromosome="16"

/map="16q23.1-q23.2"

/cell_type="high endothelial"

/tissue_type="tonsil"

1. 1992

/gene="GST3"

132. 1282

/gene="GST3"

/note="Similar to Homo sapiens sequence encoded by GenBank Accession Number AF131235"

/codon_start=1

/product="L-selectin ligand sulfotransferase GST-3"

/protein_id="AAG48246.1"

/db_xref="GI:12060808"

/translation="MLPKKMLLFLVSQMLALFPMHYNISLSMKQPERHVLVLSWRSGSSVFGQLFGQHPDVFLEPMHVMTEKQSTAMLMHMAVRDITRVLCDMSVDAFMGPRRQSLFQWENSRLCSNPACDIIPDEIIPRACRLSCQPFVEVEACRSYSHVLEKFRFNLQSLYDLKDPISLHLIHLVLRPRAVFRSRETKGLMIDSRIVMGCHQELKEEDOPYVVMVICOSELEIKTISLPRKALOEERLIVRYEDLARVPAQTSMEYFGLFELPHLOTQVHNTTRGKGMDHAFHNADALNVQSAMRMSLPREKSRQKACGDMNLGLYRHVSEQRRLDLDLSTWTPVPEIH"

BASE COUNT 482 a 540 c 489 g 481 t

ORIGIN

Query Match 84.3%; Score 1722; DB 9; Length 1992;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 1722; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 7 CATGTGTTGGTACGGAAGACGACGAAGGTTAGAGAGAAAAAGCGATGCCGCGTA 66

b 1 CATGTGTTGGTACGGAAGACGACGAAGGTTAGAGAGAAAAAGCGATGCCGCGTA 60

QY 67 GCAGTAGGCTCTCAAAAGCAGAGGAGGCCCAAGCTCTTCCACTTCAGCAC 126

Db 61 GCAGTAGGCTCTCAAAAGCAGAGGAGGCCCAAGCTCTTCCACTTCAGCAC 120

QY 127 AATGCTACTGCTCAAAAAATGAAAGCTCTGCTGTTGTTTCCAGATGCGACATCT 186

Db 122 AATGCTACTGCTCAAAAAATGAAAGCTCTGCTGTTGTTTCCAGATGCGACATCT 180

QY 187 GCGCTATTTCTTCACATGTACAGCCACAACATCACTCCCTGTCTATGAAGGACAGCC 246

Db 181 GCGCTATTTCTTCACATGTACAGCCACAACATCACTCCCTGTCTATGAAGGACAGCC 240

QY 247 CGAGGCATGCAAGCTGCTGTTCTGCTTCCTGCGGCTCTGCTCTTTTGTGGGCA 306

Db 241 CGAGGCATGCAAGCTGCTGTTCTGCTTCCTGCGGCTCTGCTCTTTTGTGGGCA 300

QY 307 GCTTTTGGGACAGCCAGATGTTTCTACCTGATGAGAGCCGCTGGAGCTGTGAT 366

Db 301 GCTTTTGGGACAGCCAGATGTTTCTACCTGATGAGAGCCGCTGGAGCTGTGAT 360

QY 367 GACCTTCAAGCAGAGCAGCGCTGATGTGTGACATGCTGTGCGGATCTGATACGGGC 426

Db 361 GACCTTCAAGCAGAGCAGCGCTGATGTGTGACATGCTGTGCGGATCTGATACGGGC 420

QY 427 CGTCTTCTGTGACATGAGCGCTTTTGTATGCTTACATGGAACCTGTGCCGAGACA 486

Db 421 CGTCTTCTGTGACATGAGCGCTTTTGTATGCTTACATGGAACCTGTGCCGAGACA 480

QY 487 GTCCAGCCCTTTCAATGAGGAGAACAGCCGGGCGCTGTCTGACACTGCGCTGATACAT 546

Db 481 GTCCAGCCCTTTCAATGAGGAGAACAGCCGGGCGCTGTCTGACACTGCGCTGATACAT 540

QY 547 CATCCCAAGATGATATATATCCCGGCGCTCACTCAGAGCTCTGTGACATCAAGCC 606

Db 541 CATCCCAAGATGATATATATCCCGGCGCTCACTCAGAGCTCTGTGACATCAAGCC 600

QY 607 CTTTGAAGTGTGAGAGAGCGCTGCCCTTCAACAGCCACAGTGTGCTCAAGAGGTGCG 666

Db 601 CTTTGAAGTGTGAGAGAGCGCTGCCCTTCAACAGCCACAGTGTGCTCAAGAGGTGCG 660

QY 667 CTTCTTCAACCTGCACTCCCTCTACCGCGCTGAGAAAGCCCTCCCTCACTGATATAT 726

Db 661 CTTCTTCAACCTGCACTCCCTCTACCGCGCTGAGAAAGCCCTCCCTCACTGATATAT 720

QY 727 CGTGCACTGTGTCGGGAGACCCCGGCGCTGTCCGTTCCCGAAGACGACAAAGGAGA 786

Db 721 CGTGCACTGTGTCGGGAGACCCCGGCGCTGTCCGTTCCCGAAGACGACAAAGGAGA 780

QY 787 TCTCATGATTGACAGTGCATTTGTATGGGCGACATGACAAAAAATCAAGAGAGGA 846

Db 781 TCTCATGATTGACAGTGCATTTGTATGGGCGACATGACAAAAAATCAAGAGAGGA 840

QY 847 CCAACCTCTATGATGATGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 906

Db 841 CCAACCTCTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 900

QY 907 CCAAGCTTGTCCCAAGGCGCTGAGAGACGTAACGCTTGTGCGTATGAGAGCTGCG 966

Db 901 CCAAGCTTGTCCCAAGGCGCTGAGAGACGTAACGCTTGTGCGTATGAGAGCTGCG 960

QY 967 TCGAGCCCTGTGCGGCGCAACCTCCGAATGATGATGATGATGATGATGATGATGATGAT 1026

Db 961 TCGAGCCCTGTGCGGCGCAACCTCCGAATGATGATGATGATGATGATGATGATGATGAT 1020

QY 1027 CCATCTTCAAGACCTGGGTGATATACATCAACCCGAGCAAGGCGATGGTGACACGCTTT 1086

Db 1021 CCATCTTCAAGACCTGGGTGATATACATCAACCCGAGCAAGGCGATGGTGACACGCTTT 1080

QY 1087 CCAACAAATGCCAGGATGCCCTTAATGTCCTCCAGGCTTGGCGTCTTGGCCTTA 1146

Db 1081 CCAACAAATGCCAGGATGCCCTTAATGTCCTCCAGGCTTGGCGTCTTGGCCTTA 1140

QY 1147 TGAAGGTTTTCGACTTCAGAAAGCTGTGCGATGCGAATGATGATGATGATGATGATGATGAT 1206

Db 1141 TGAAGGTTTTCGACTTCAGAAAGCTGTGCGATGCGAATGATGATGATGATGATGATGATGAT 1200

QY 1207 CCAAGCTCAGATCGAACAAGAACAGAAAACTGTGTGATGATCTTCTCTCACTGAGAC 1266

Db 1201 CCAAGCTCAGATCGAACAAGAACAGAAAACTGTGTGATGATCTTCTCTCACTGAGAC 1260

QY 1267 TGTCCCTGAGCAAAATCACTAAGAGGTTGAGAAAGCTTGTGCTCCACTGTGTACGCC 1326

Db 1261 TGTCCCTGAGCAAAATCACTAAGAGGTTGAGAAAGCTTGTGCTCCACTGTGTACGCC 1320

QY 1327 TCAAGTACCTTCTGTAATGCTCTGAGGCTTCCGACAACTCTGAGGCTTAACTCACTACATG 1386

Db 1321 TCAAGTACCTTCTGTAATGCTCTGAGGCTTCCGACAACTCTGAGGCTTAACTCACTACATG 1380

QY 1387 TCTGTGGGTATCACTAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1446

Db 1381 TCTGTGGGTATCACTAGATGTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1440

QY 1447 GTGTCCATGCTTGTGTCTAGAAAAACAGACTGGGGAACCTTATGTGAGACACATCCAC 1506

Db 1441 GTGTCCATGCTTGTGTCTAGAAAAACAGACTGGGGAACCTTATGTGAGACACATCCAC 1500

QY 1507 CAGTGAACAGGATATGCTCTTCTTCTTCTGATCTTCTCTGCTGCTGCTGCTGCTGCTGCTGCT 1566


```

QY 951 GCTATAGAGACCTGGCTCGAGCCCTGTGGCCAGACTTCCGATGATGAATTCGTGG 1010
Db 901 GCTATAGAGACCTGGCTCGAGCCCTGTGGCCAGACTTCCGATGATGAATTCGTGG 960
QY 1011 GATTGGAATTCCTGGCCCATCTTCAGACCTGGGTGCATACATCACCAGGAGAGGCA 1070
Db 961 GATTGGAATTCCTGGCCCATCTTCAGACCTGGGTGCATACATCACCAGGAGAGGCA 1020
QY 1071 TGGGTGACCAAGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGC 1130
Db 1021 TGGGTGACCAAGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCTTGGC 1080
QY 1131 GCGGCTCTTGGCCCTTGAAGAGCTTCGATTCAGAAAAGCTGTGGGATGCCATGA 1190
Db 1081 GCGGCTCTTGGCCCTTGAAGAGCTTCGATTCAGAAAAGCTGTGGGATGCCATGA 1140
QY 1191 ATTGCTGGGCTACCCGACAGTCAGATCGAACAAGAGAGAACTTGTCTGCTGATC 1250
Db 1141 ATTGCTGGGCTACCCGACAGTCAGATCGAACAAGAGAACTTGTCTGCTGATC 1200
QY 1251 TTCTGCTACCTGGAGCTGTCCCTGAGCAAAATCCATTAAGAGGCTTGAAGAGCTTGTG 1310
Db 1201 TTCTGCTACCTGGAGCTGTCCCTGAGCAAAATCCATTAAGAGGCTTGAAGAGCTTGTG 1260
QY 1311 CCACCTGGTGTGAGCTGACATCTTCTGTAATGCTTGTGAGCTTGTGCTACATCTCT 1370
Db 1261 CCACCTGGTGTGAGCTGACATCTTCTGTAATGCTTGTGAGCTTGTGCTACATCTCT 1320
QY 1371 GAGCCTTAACATACATGTCTGTGGGTATCAACACTGATGTGAGTGTGTCACACAGTCTC 1430
Db 1321 GAGCCTTAACATACATGTCTGTGGGTATCAACACTGATGTGAGTGTGTCACACAGTCTC 1380
QY 1431 AAGCAGAAGAGCTTTTGTGTCCATGCTTGTGTAGAAAACAGACTGGGAACTTATGT 1490
Db 1381 AAGCAGAAGAGCTTTTGTGTCCATGCTTGTGTAGAAAACAGACTGGGAACTTATGT 1440
QY 1491 GAGCAGACATCCACAGCAAGGAAAGAGGATGCTTCTTCTTGTGATCTCTCCG 1550
Db 1441 GAGCAGACATCCACAGCAAGGAAAGAGGATGCTTCTTCTTGTGATCTCTCCG 1500
QY 1551 TCTGGGACAGACTTCAGAGACTTGTGGCTGTGAGGAGCTTAAAGCAGACAGATATCAG 1610
Db 1501 TCTGGGACAGACTTCAGAGACTTGTGGCTGTGAGGAGCTTAAAGCAGACAGATATCAG 1560
QY 1611 TGGATTTGATCCATTAACCTCCCTGTCCACATCTTGGCCCAATGGGAATGATCTTTCAC 1670
Db 1561 TGGATTTGATCCATTAACCTCCCTGTCCACATCTTGGCCCAATGGGAATGATCTTTCAC 1620
QY 1671 CAAGAGCTCAGCAGACTTTTCCACAGAGATGC 1703
Db 1621 CAAGAGCTCAGCAGACTTTTCCACAGAGATGC 1653

RESULT 3
AK026635 2011 bp mRNA linear pri 29-sep-2000
LOCUS Homo sapiens cDNA: FLJ22982 f1s, clone KAT11454, highly similar to
DEFINITION AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase
mRNA.
ACCESSION AK026635
VERSION AK026635.1 GI:10439531
KEYWORDS oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
1 (sites)
REFERENCE Matanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H.,
AUTHORS Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T.,
Nakamura,Y., Isogai,T. and Sugano,S.
TITLE NED0 human cDNA sequencing project
JOURNAL NEBO human cDNA sequencing project
REFERENCE Unpublished
2 (bases 1 to 2011)

```

```

AUTHORS Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T.,
TITLE Shibahara,T., Tanaka,T. and Nakamura,Y.
JOURNAL Direct Submission
Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science,
University of Tokyo, Laboratory of Genome Structure Analysis, Human
Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639,
Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286,
Fax:81-3-5449-5416)
COMMENT NED0 human cDNA sequencing project supported by Ministry of
International Trade and Industry of Japan: cDNA full insert
sequencing: Research Association for Biotechnology: cDNA library
construction, 5'-3'-end one pass sequencing: Department of
Virology and Human Genome Center, Institute of Medical Science,
University of Tokyo (partly supported by Science and Technology
Agency).
FEATURES
source location/Qualifiers
1..2011
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="KAT11454"
/cell_line="KATO III"
/cell_type="signed-rising cell carcinoma"
/clone_idb="KAT"
/notes="cloning vector pME18SFL3"
1..2011
/notes="highly similar to AF131235 Homo sapiens
N-acetylglucosamine 6-O-sulfotransferase mRNA"
BASE COUNT 482 a 541 c 475 g 513 t
ORIGIN
Query Match 78.2%; Score 1598; DB 9; Length 2011;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1598; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 106 CAAGGCTTCCACTTACGACAATGCTACTGCTTAAATGAAGAACTCTGCTGTCT 165
Db 50 CAAGGCTTCCACTTACGACAATGCTACTGCTTAAATGAAGAACTCTGCTGTCT 109
QY 166 GGTTCACAGAGGAGCATCTTGGCTATCTTCCACATGATCAGCACAATCAGCTC 225
Db 110 GGTTCACAGAGGAGCATCTTGGCTATCTTCCACATGATCAGCACAATCAGCTC 169
QY 226 CCTGTATTAAGAGCAGACCCGAGCGATGCAAGTGTGTCTCTCTGCGGCTC 285
Db 170 CCTGTATTAAGAGCAGACCCGAGCGATGCAAGTGTGTCTCTCTGCGGCTC 229
QY 286 TGGCTCTCTTTGTGGGAGCGCTTTTGGGAGCAGCAGAGTCTTCTACCTGATGGA 345
Db 230 TGGCTCTCTTTGTGGGAGCGCTTTTGGGAGCAGCAGAGTCTTCTACCTGATGGA 289
QY 346 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGACCGCTGATCTGCACATGCG 405
Db 290 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGACCGCTGATCTGCACATGCG 349
QY 406 TGTGGGAGTCTGATACGCGCGCTCTTGTGTGCGACATGAGCGCTTGTGATGCTACT 465
Db 350 TGTGGGAGTCTGATACGCGCGCTCTTGTGTGCGACATGAGCGCTTGTGATGCTACT 409
QY 466 GGACCTGTGCTCCGAGAGACGACGCTCTTCAAGTGGGAGAACAGCGCGGCGCTGG 525
Db 410 GGACCTGTGCTCCGAGAGACGACGCTCTTCAAGTGGGAGAACAGCGCGGCGCTGG 469
QY 526 TTCTGACCTGCTGTGACATATCCACAAAGATGAATCATCCCGGCGTCACTGACG 585
Db 470 TTCTGACCTGCTGTGACATATCCACAAAGATGAATCATCCCGGCGTCACTGACG 529
QY 586 GCTCTGTGACATCAACAGCCCTTGAAGTGTGTGAAGAGCGCTCGCTCTTACAGCA 645
Db 530 GCTCTGTGACATCAACAGCCCTTGAAGTGTGTGAAGAGCGCTCGCTCTTACAGCA 589
QY 646 CGTGTGCTCAAGAGGTCGCTTCTTCAACTGCAAGTCCCTTACCGGCTCTGAAGA 705

```

Db 590 CGTGTCCTCAGAGAGGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTGAAGA 649
 QY 706 CCCCTCCCTCAACCTGATATCGNGACCTGTCGGGAGCCCCGGGCGCTTCCCTTC 765
 Db 650 CCCCTCCCTCAACCTGATATCGNGACCTGTCGGGAGCCCCGGGCGCTTCCCTTC 709
 QY 766 CCGAGAACGACAAAGGAGATCTCATGATTTGACAGTGCATTTGTATGGGAGCATGA 825
 Db 710 CCGAGAACGACAAAGGAGATCTCATGATTTGACAGTGCATTTGTATGGGAGCATGA 769
 QY 826 GCAAAAACCTCAAGAGAGAGACCAACCTTATGTATGATGATCATCTCCAAAGCA 885
 Db 770 GCAAAAACCTCAAGAGAGAGACCAACCTTATGTATGATGATGATCATCTCCAAAGCA 829
 QY 886 GCTGAGATCTACAAGACCATCCAGTCTTGCCCAAGGCCCTGAGAGAGCTTCCCTGCT 945
 Db 830 GCTGAGATCTACAAGACCATCCAGTCTTGCCCAAGGCCCTGAGAGAGCTTCCCTGCT 889
 QY 946 TGTGCGCTATGAGAGACCTGGCTGAGGCCCTGTGGCCAGACTTCCGGAATGTATGAAT 1005
 Db 890 TGTGCGCTATGAGAGACCTGGCTGAGGCCCTGTGGCCAGACTTCCGGAATGTATGAAT 949
 QY 1006 CGTGGGATTTGGAATTTCTTGGCCCATCTTTCAGAGACTGGGTGATTAATCACCCGAGCAA 1065
 Db 950 CGTGGGATTTGGAATTTCTTGGCCCATCTTTCAGAGACTGGGTGATTAATCACCCGAGCAA 1009
 QY 1066 GGGCATGGTGAACCAAGCTTTCACACAAATGCCAGGAGTGGCCCTTAATGTCTCCAGGC 1125
 Db 1010 GGGCATGGTGAACCAAGCTTTCACACAAATGCCAGGAGTGGCCCTTAATGTCTCCAGGC 1069
 QY 1126 TTGGCGCTGCTTCTTGGCCCATATGAAGAGTTTCTGCACTTCAGAAAGCTGTGGCGATGC 1185
 Db 1070 TTGGCGCTGCTTCTTGGCCCATATGAAGAGTTTCTGCACTTCAGAAAGCTGTGGCGATGC 1129
 QY 1186 CATGAATTTCTTGGCCCATACCCGCGCATCATCTGAACACAGAGAAACCTGTGCT 1245
 Db 1130 CATGAATTTCTTGGCCCATACCCGCGCATCATCTGAACACAGAGAAACCTGTGCT 1189
 QY 1246 GGAATCTTCTGCTACCTGAGTGCCTGAGCAATCCACTAAGAGGTTGAGAAAGCTT 1305
 Db 1190 GGAATCTTCTGCTACCTGAGTGCCTGAGCAATCCACTAAGAGGTTGAGAAAGCTT 1249
 QY 1306 TGTGCGCACTGCTGTGAGGCTCAGTCACTTCTGTAATGCTTCTGAGCCCTTGCCTACA 1365
 Db 1250 TGTGCGCACTGCTGTGAGGCTCAGTCACTTCTGTAATGCTTCTGAGCCCTTGCCTACA 1309
 QY 1366 TCTCTGAGCCTTAATCAATGATCTGTGGGATTCACACTGATGATGTGTCCACACG 1425
 Db 1310 TCTCTGAGCCTTAATCAATGATCTGTGGGATTCACACTGATGATGTGTCCACACG 1369
 QY 1426 TGTCTCAAGCAGAGAGACTTTTGTGTCCATGCTTGTCTGTAAGAAAGAGACTGGGAACT 1485
 Db 1370 TGTCTCAAGCAGAGAGACTTTTGTGTCCATGCTTGTGTCTGTAAGAAAGAGACTGGGAACT 1429
 QY 1486 TATGTGAGCAGACATCCCAACCAAGTGAAGAGGGTATGCTCTTCTTCTTCTTGTGATCT 1545
 Db 1430 TATGTGAGCAGACATCCCAACCAAGTGAAGAGGGTATGCTCTTCTTCTTCTTGTGATCT 1489
 QY 1546 TCTCTGCTGGGAGAGACTTTCAGAGACTTTGTGGCTTGAAGGCTTATTAAGCAGCAGACAGT 1605
 Db 1490 TCTCTGCTGGGAGAGACTTTCAGAGACTTTGTGGCTTGAAGGCTTATTAAGCAGCAGACAGT 1549
 QY 1606 ATCAGTGAATTTGATCCATAAACCCTCCCTGTCCACATCTTGTCCCAATGGGAAATGATCT 1665
 Db 1550 ATCAGTGAATTTGATCCATAAACCCTCCCTGTCCACATCTTGTCCCAATGGGAAATGATCT 1609
 QY 1666 TTCAACCAAGAGCTCACACAGATTTTCCACAGAGATGC 1703
 Db 1610 TTCAACCAAGAGCTCACACAGATTTTCCACAGAGATGC 1647

RESULT 4
 AC010547

LOCUS AC010547 183228 bp DNA linear PRI 29-SEP-2001
 DEFINITION Homo sapiens chromosome 16 clone RP11-510M2, complete sequence.
 ACCESSION AC010547
 VERSION AC010547.9 GI:15808510
 KEYWORDS HTG.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE Direct Submission
 JOURNAL Submitted (29-SEP-2001) Production Sequencing Facility, DOE Joint
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 COMMENT
 Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 Genbank flat file format but are available as part
 of this entry's ASN.1 file.

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality
 estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than
 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the
 Genbank flat file format but are available as part
 of this entry's ASN.1 file.

 FEATURES
 source
 1. 183228
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-510M2"
 BASE COUNT 51901 a 42098 c 40026 g 49203 t
 ORIGIN
 Query Match 78.2%; Score 1597; DB 9; Length 183228;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Db 107 AAGGTCTTCCACTTCACGACATGCTACTGCTTAAAAAATGAAGAGCTTCGCTTTCTG 166
 QY 167 GTTTCACAGATGGGCACTTGGCTATTTCTTCCAGATGTACAGCAGCAACATAGAGCTCC 226
 Db 142584 AAGGTCTTCCACTTCACGACATGCTACTGCTTAAAAAATGAAGAGCTTCGCTTTCTG 142643
 QY 142844 GTTTCACAGATGGGCACTTGGCTATTTCTTCCAGATGTACAGCAGCAACATAGAGCTCC 142703
 Db 227 CTGTCTATGAAGGACGACGCCGAGCATGACGCTGTGTTCTTCTTCTGCGCTCT 286
 QY 142704 CTGTCTATGAAGGACGACGCCGAGCATGACGCTGTGTTCTTCTTCTGCGCTCT 142763
 Db 287 GGCCTCTCTTTTGTGGGGAGACTTTTGGGAGCAGCCAGATGTTTCTACCTGATGAG 346
 QY 142764 GGCCTCTCTTTTGTGGGGAGACTTTTGGGAGCAGCCAGATGTTTCTACCTGATGAG 142823

OY	347	CCCGGCTGGCAGTGTGGATGACATTCACAAAGCAGACACCGCGTGGATGCTGCACATAGCCT	142824
Db	142824	CCCGGCTGGCAGTGTGGATGACATTCACAAAGCAGACACCGCGTGGATGCTGCACATAGCCT	142883
OY	407	GTGGGGGATCTGATACGGGGCGCTCTTCTTGTGCGACATGAGCGCTTTGATGCTTACATG	142884
Db	142884	GTGGGGGATCTGATACGGGGCGCTCTTCTTGTGCGACATGAGCGCTTTGATGCTTACATG	142943
OY	467	GAACCTGGTCCCGGAGACAGTCACAGCCCTTCATGGTGGAAACAACAGCCGGGCGCTGT	142944
Db	142944	GAACCTGGTCCCGGAGACAGTCACAGCCCTTCATGGTGGAAACAACAGCCGGGCGCTGT	143003
OY	527	TCTGCACCTGCGCTGTGACATCATCCACAAAGATGAATCATCCCCGGGCTCACTGCAAG	143004
Db	143004	TCTGCACCTGCGCTGTGACATCATCCACAAAGATGAATCATCCCCGGGCTCACTGCAAG	143063
OY	587	CTCCCTGTGCAATCAACAGCCCTTTAGAGTGGTGGAGAAAGGCTGCCGCTCTTACAGCCAC	143064
Db	143064	CTCCCTGTGCAATCAACAGCCCTTTAGAGTGGTGGAGAAAGGCTGCCGCTCTTACAGCCAC	143123
OY	647	GTGGGCTCAAGAGAGGTGGCTTCTTCAACCTGACAGTCCCTTACCCGGTGGTGAAGAG	143124
Db	143124	GTGGGCTCAAGAGAGGTGGCTTCTTCAACCTGACAGTCCCTTACCCGGTGGTGAAGAG	143183
OY	707	CCCTCCCTCAACCTGCAATATCGTGCACCTGTGTCGGGAGCCCGGGCGCTGTTCGTTCC	143184
Db	143184	CCCTCCCTCAACCTGCAATATCGTGCACCTGTGTCGGGAGCCCGGGCGCTGTTCGTTCC	143243
OY	767	CGAGAGCGCAAAAGGGAGATCTCAATGATGACATCGCATTTGTGATGGGGGACAGCATAG	143244
Db	143244	CGAGAGCGCAAAAGGGAGATCTCAATGATGACATCGCATTTGTGATGGGGGACAGCATAG	143303
OY	827	CAAAAACTCAGAAAGAGAGACCACACCTCACTATGTGATGACAGTCACTGCCAAAGCCAG	143304
Db	143304	CAAAAACTCAGAAAGAGAGACCACACCTCACTATGTGATGACAGTCACTGCCAAAGCCAG	143363
OY	887	CTGGGATCTTACAGACCATTCACAGTCTTGCCCAAGGCCCTTCAGAAAGCTTACTGCTT	143364
Db	143364	CTGGGATCTTACAGACCATTCACAGTCTTGCCCAAGGCCCTTCAGAAAGCTTACTGCTT	143423
OY	947	GTGCGCTATGAGAGACCTGGCTGACAGCCCTGTGGGCCAGACTTCCGATGTATGAATTC	143424
Db	143424	GTGCGCTATGAGAGACCTGGCTGACAGCCCTGTGGGCCAGACTTCCGATGTATGAATTC	143483
OY	1007	GTGGGATTTGAATTTTGGCCCATCTTACAGACCTGGGTGCATTAATCATCCCGAGGCAAG	143484
Db	143484	GTGGGATTTGAATTTTGGCCCATCTTACAGACCTGGGTGCATTAATCATCCCGAGGCAAG	143543
OY	1067	GGCATGGGTGACACAGCTTTCACACAATGACAGAGGATGCCCTTAATGTCCCGCAGGT	143544
Db	143544	GGCATGGGTGACACAGCTTTCACACAATGACAGAGGATGCCCTTAATGTCCCGCAGGT	143603
OY	1127	TGGCGCTGCTTTTGGCCCATATGAAAGTTTCTCGACTTTCAGAAAGCCGTTGGCGATGCC	143604
Db	143604	TGGCGCTGCTTTTGGCCCATATGAAAGTTTCTCGACTTTCAGAAAGCCGTTGGCGATGCC	143663
OY	1187	ATGAATTTGCTGGGCTTACCGCCAGCTCAGATGTGAACAAGAACAGAAACCTGTTCGTG	143664
Db	143664	ATGAATTTGCTGGGCTTACCGCCAGCTCAGATGTGAACAAGAACAGAAACCTGTTCGTG	143723
OY	1247	GATCTTCTGTACTACCTGAGACTGTCCCTAGACAAATCCACTAAGAGGTTGAAGAGCTTT	143724
Db	143724	GATCTTCTGTACTACCTGAGACTGTCCCTAGACAAATCCACTAAGAGGTTGAAGAGCTTT	143783
OY	1307	GCTGCGACCTGGTGTCAAGCTTCACTACTTCTCTGAATGCTTGTGAGCCCTTGCTACAT	143784
Db	143784	GCTGCGACCTGGTGTCAAGCTTCACTACTTCTCTGAATGCTTGTGAGCCCTTGCTACAT	143843
OY	1367	CTGTGAGCCTTAACATACATGCTGTGGGATATACACTGATGCTGATGTGTGCCACAGCT	143844
Db	143844	CTGTGAGCCTTAACATACATGCTGTGGGATATACACTGATGCTGATGTGTGCCACAGCT	143903

OY	1427	GCTCAGACGAGAAGCACTTTTGTGCATCAGCTTGTCATGAAGAAACAGACTGGGAACTT	1486
Db	143904	GCTCAGACGAGAAGCACTTTTGTGCATCAGCTTGTCATGAAGAAACAGACTGGGAACTT	1439
OY	1487	ATGTAGACGACACATCCACCACAGTAAGAGGATTGCTTCTTTCTTTCTTGATCTT	1546
Db	143964	ATGTAGACGACACATCCACCACAGTAAGAGGATTGCTTCTTTCTTTCTTGATCTT	1440
OY	1547	CCTGCTCGGGAGACTTAGAGACTTTTGCGCTGGAGAGCTATTAAACAGCACAGTA	1606
Db	144024	CCTGCTCGGGAGACTTAGAGACTTTTGCGCTGGAGAGCTATTAAACAGCACAGTA	1440
OY	1607	TCAGTGAATTGATTCATATAAACCCTCCCTGTCCACATCTTGCCCATGGGGAGATCTT	1666
Db	144084	TCAGTGAATTGATTCATATAAACCCTCCCTGTCCACATCTTGCCCATGGGGAGATCTT	1441
OY	1667	TCACCAAAGAGCTCACCGACATTTTCCACACAGATGC	1703
Db	144144	TCACCAAAGAGCTCACCGACATTTTCCACACAGATGC	144180
<hr/>			
RESULT 5	AC138848/c	206943 bp DNA linear HTG 21-JAN-2003	
LOCUS	AC138848	Homo sapiens chromosome 5 clone RP11-1301B21, WORKING DRAFT	
DEFINITION	SEQUENCE, 7 unordered pieces.		
ACCESSION	AC138848		
VERSION	AC138848.1 GI:27805260		
KEYWORDS	HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.		
SOURCE	Homo sapiens (human)		
ORGANISM	Homo sapiens		
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;		
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.		
TITLE	1 (bases 1 to 206943)		
JOURNAL	DOE Joint Genome Institute.		
REFERENCE	Sequencing of Human Chromosome 5		
AUTHORS	unpublished		
TITLE	2 (bases 1 to 206943)		
JOURNAL	DOE Joint Genome Institute.		
REFERENCE	Direct Submission		
AUTHORS	Submitted (21-JAN-2003) Production Sequencing Facility, DOE Joint		
TITLE	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA		
JOURNAL	-----Genome Center		
COMMENT	Center: Joint Genome Institute		
	Center Code: JGI		
	Web site: http://www.jgi.doe.gov		

	Project Information		
	Center Project Name: 2719263		
	Center clone name: RPCI-1L_1301B21		

	Summary Statistics		
	Consensus quality: 202039 bases at least Q40		
	Consensus quality: 202804 bases at least Q30		
	Consensus quality: 203337 bases at least Q20		
	Estimated insert size: 175000; agarose-fp estimation		
	Estimated insert size: 206343; sum-of-contigs estimation		
	Quality coverage: 15.28 in Q20 bases; agarose-fp estimation		
	Quality coverage: 12.96 in Q20 bases; sum-of-contigs estimation.		
	* NOTE: This is a 'working draft' sequence. It currently		
	* consists of 7 contigs. The true order of the pieces		
	* is not known and their order in this sequence record is		
	* arbitrary. Gaps between the contigs are represented as		
	* runs of N, but the exact sizes of the gaps are unknown.		
	* This record will be updated with the finished sequence		
	* as soon as it is available and the accession number will		
	* be preserved.		
*	1	1219: contig of 1219 bp in length	
*	1220	1319: gap of unknown length	
*	1320	2493: contig of 1174 bp in length	
*	2494	2593: gap of unknown length	
*	2594	3666: contig of 1073 bp in length	
*	3667	3766: gap of unknown length	

```

* 3767 5167: contig of 1401 bp in length
* 5168 5267: gap of unknown length
* 5268 12818: contig of 7551 bp in length
* 12819 12918: gap of unknown length
* 12919 36169: contig of 23251 bp in length
* 36170 36269: gap of unknown length
* 36270 206943: contig of 170674 bp in length.
Location/Qualifiers
1. 206943
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RP11-1301B21"
/clone_lib="RP11 human BAC library 11"
BASE COUNT 56312 a 44122 c 46163 g 59225 t 1121 others
ORIGIN
Query Match 78.2% Score 1597; DB 2; Length 206943;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Y 107 AAGGTCCTCCACTTCACGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTTTCTG 166
|||||
125368 AAGGTCCTCCACTTCACGACAAATGCTACTGCTTAAAAAATGAAGCTCTGCTTTCTG 125309
|||||
167 GTTCCAGATGGCATCTTGGCTCTATTTCTCCACATGTACAGCAACATCATGAGCTCC 226
|||||
125308 GTTCCAGATGGCATCTTGGCTCTATTTCTCCACATGTACAGCAACATCATGAGCTCC 125249
|||||
227 CTGTCTATGAAGGACAGACCCGAGCGATGACAGCTGCTGTCTCTCTCTGCGCTCT 286
|||||
125248 CTGTCTATGAAGGACAGACCCGAGCGATGACAGCTGCTGTCTCTCTCTGCGCTCT 125189
|||||
287 GGCTCTCTTTTGGGGGACCTTTTGGGAGACACCAATGTTTCTTACTGATGAG 346
|||||
125188 GGCTCTCTTTTGGGGGACCTTTTGGGAGACACCAATGTTTCTTACTGATGAG 125129
|||||
347 CCCGCTGCGACGCTGTGATGACCTTCAAGAGAGACACCGCTGTGATGCTGATGCT 406
|||||
125128 CCCGCTGCGACGCTGTGATGACCTTCAAGAGAGACACCGCTGTGATGCTGATGCT 125069
|||||
407 GTGCGGATCTGATACGCGCGCTTCTTGTGAGACATGACGCTTGTGATGCTGATGCT 466
|||||
125068 GTGCGGATCTGATACGCGCGCTTCTTGTGAGACATGACGCTTGTGATGCTGATGCT 125009
|||||
467 GAACCTGTGCTCCGAGACAGCTCAAGCTCTTCAAGTGGAGAAACGCGCGCTGTGT 526
|||||
125008 GAACCTGTGCTCCGAGACAGCTCAAGCTCTTCAAGTGGAGAAACGCGCGCTGTGT 124949
|||||
527 TCTGACCTGCTGCTGATCATGCTCAACAAAGATCAATCCCGCGCTGCTGACGAG 586
|||||
124948 TCTGACCTGCTGCTGATCATGCTCAACAAAGATCAATCCCGCGCTGCTGACGAG 124889
|||||
587 CTCCTGTGAGTCAACAGCCCTTGTGAGTGTGAGAGAGCTCCGCTCTCTACAGCCAC 646
|||||
124888 CTCCTGTGAGTCAACAGCCCTTGTGAGTGTGAGAGAGCTCCGCTCTCTACAGCCAC 124829
|||||
647 GTGCTGCTCAAGAGAGTGGCTTCTTCAACCTGACGTCCCTCAACCGCTGTGAAGAC 706
|||||
124828 GTGCTGCTCAAGAGAGTGGCTTCTTCAACCTGACGTCCCTCAACCGCTGTGAAGAC 124769
|||||
707 CCCCTCCCAACCGCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 766
|||||
124768 CCCCTCCCAACCGCATATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 124709
|||||
767 CGAAGAGCAACAAAGAGATCTCATGATTGACAGTGCATTTGATGGGAGCAGATGAG 826
|||||
124708 CGAAGAGCAACAAAGAGATCTCATGATTGACAGTGCATTTGATGGGAGCAGATGAG 124649
|||||
827 CAAAACCTCAAGAGAGAGCAACCTTACTATGTGATGACAGTGCATCTGCAAGAGCCAG 886
|||||
124648 CAAAACCTCAAGAGAGAGCAACCTTACTATGTGATGACAGTGCATCTGCAAGAGCCAG 124589
|||||

```

```

OY 887 CTGAGATCTTCAAGAGACATCCAGTCCCTTGGCCAAAGGCCCTGAGAAAGCTTACTGCTT 946
|||||
124588 CTGAGATCTTCAAGAGACATCCAGTCCCTTGGCCAAAGGCCCTGAGAAAGCTTACTGCTT 124529
|||||
OY 947 GTGCGCTATGAGAGACCTGGCTGAGAGCCCTTGGGCCAGACTTCCGAAATGAAATTC 1006
|||||
124528 GTGCGCTATGAGAGACCTGGCTGAGAGCCCTTGGGCCAGACTTCCGAAATGAAATTC 124469
|||||
OY 1007 GTGGATTTGGAATTTCTTCCCATCTTCAACCTTGGTGTGCTTAACATCACCAGGCAAG 1066
|||||
124468 GTGGATTTGGAATTTCTTCCCATCTTCAACCTTGGTGTGCTTAACATCACCAGGCAAG 124409
|||||
OY 1067 GGCATGGGTGACACAGCTTTCACACAAATGCGAGGATGGCCCTTAATGTCTCCAGGCT 1126
|||||
124408 GGCATGGGTGACACAGCTTTCACACAAATGCGAGGATGGCCCTTAATGTCTCCAGGCT 124349
|||||
OY 1127 TGGCGCTGCTTTCCTTATGAAAGGCTTCTGACTTCAGAAAGCCTGTGGGATGCC 1186
|||||
124348 TGGCGCTGCTTTCCTTATGAAAGGCTTCTGACTTCAGAAAGCCTGTGGGATGCC 124289
|||||
OY 1187 ATGAATTTGCTGGGCTACCGGACGTGATCTGAAGAAAGAGAAACCTGTGCTG 1246
|||||
124288 ATGAATTTGCTGGGCTACCGGACGTGATCTGAAGAAAGAGAAACCTGTGCTG 124229
|||||
OY 1247 GATCTTCTTACTCTGACTGTCCCTGAGCAAAATCCCTTAAGAGGTTGAGAGGCTTT 1306
|||||
124228 GATCTTCTTACTCTGACTGTCCCTGAGCAAAATCCCTTAAGAGGTTGAGAGGCTTT 124169
|||||
OY 1307 GCTGCCACCTGCTGACAGCTGACAGTCACTTCTCTGAATGCTTGTGAGCCTTGCCTACAT 1366
|||||
124168 GCTGCCACCTGCTGACAGCTGACAGTCACTTCTCTGAATGCTTGTGAGCCTTGCCTACAT 124109
|||||
OY 1367 CTCTGAGCTTAACTACATGCTGTGGGTATCACACTGATGATGATGATGATGATGATGAT 1426
|||||
124108 CTCTGAGCTTAACTACATGCTGTGGGTATCACACTGATGATGATGATGATGATGATGAT 124049
|||||
OY 1427 GCTCAAGAGAGAGAGACTTGTGTCACAGCTTGTGTGAGAAACAGAGCTGGGAACTT 1486
|||||
124048 GCTCAAGAGAGAGAGACTTGTGTCACAGCTTGTGTGAGAAACAGAGCTGGGAACTT 123989
|||||
OY 1487 ATGTGAGCAGACATCCACAGTGAAGAGGATGCTCTTCTTCTTCTGATCTT 1546
|||||
123988 ATGTGAGCAGACATCCACAGTGAAGAGGATGCTCTTCTTCTTCTGATCTT 123929
|||||
OY 1547 CCTGTCTGGGAGACTTGAAGACTTGTGGCTGTGAGAGCCTATTAAACAGACACACTA 1606
|||||
123928 CCTGTCTGGGAGACTTGAAGACTTGTGGCTGTGAGAGCCTATTAAACAGACACACTA 123869
|||||
OY 1607 TCAGTGAATGATGATCAATTAACCTCCCTGTCACATCTTGGCCAAATGGGAAATGATCTT 1666
|||||
123868 TCAGTGAATGATGATCAATTAACCTCCCTGTCACATCTTGGCCAAATGGGAAATGATCTT 123809
|||||
OY 1667 TCACCAAGAGACTCACAGACTTTCACAGAGATGC 1703
|||||
123808 TCACCAAGAGACTCACAGACTTTCACAGAGATGC 123772
|||||
Db 123808 TCACCAAGAGACTCACAGACTTTCACAGAGATGC 123772
|||||

```

RESULT 6
 LOCUS BD127258 1979 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD127258.1 GI:23222203
 VERSION BD127258.1
 KEYWORDS JP 2002017375-A/2689.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1979)
 Ota, T., Nishikawa, T., Isogai, T., Hayashi, K., Ishii, S., Kawai, Y.,
 Makamatsu, A., Sugiyama, T., Nagai, K., Kojima, S., Otsuki, T. and
 Koga, H.

VERSION AK074746.1 GI:22760388
KEYWORDS Oligo capping; f1s (full insert sequence).
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1
AUTHORS Isogai,T., Ota,T., Nishikawa,T., Hayashi,K., Otsuki,T.,
Sugiyama,T., Suzuki,Y., Nagai,K., Sugano,S., Ishii,S.,
Kawai-Hio,Y., Saito,K., Yamamoto,J., Wakamatsu,A., Nakamura,Y.,
Kojima,S., Nagahara,K., Masuhara,Y., Ono,T., Okano,K., Yoshikawa,Y.,
Aotsuka,S., Sasaki,N., Hattori,A., Okumura,K., Iwayanagi,T. and
Mihomiyu,K.
TITLE NEDO human cDNA sequencing project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 1979)
AUTHORS Isogai,T. and Otsuki,T.
TITLE Direct Submission
JOURNAL Submitted (25-MAR-2002) Takao Isogai, Helix Research Institute,
Genomics Laboratory, 1532-3 Yana, Kisarazu, Chiba 292-0812, Japan
(E-mail:genomics@hri.co.jp, Tel:81-438-52-3975, Fax:81-438-52-3986)
COMMENT NEDO human cDNA sequencing project supported by Ministry of
Economy, Trade and Industry of Japan; cDNA full insert sequencing:
Research Association for Biotechnology; cDNA library construction:
Institute of Medical Science, University of Tokyo, Laboratory of
Genome Structure, Human Genome Center; cDNA 5'- & 3'-end one pass
sequencing and clone selection: Helix Research Institute (supported
by Japan Key Technology Center etc.).
FEATURES
Source
1. 1979
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="NF2RM4001325"
/cell_line="NT2"
/cell_type="teratocarcinoma"
/clone_id="NF2RM4"
/note="cloning vector: PME18SFL3-mRNA from uninduced NT2
neuronal precursor cells"
69..1329
/note="unamed protein product"
/codon_start=1
/protein_id="BAC11177.1"
/db_xref="GI:22760389"
/translation="MLPKKMKLLPLVSOMALLPFFHMYSHNINSLSMKOPBRMH
VIVLSRSGSSFYGLKPGHPDVFYLMERPAHYHMTFKOSTAMHLHAYVDLIRAY
LCHDSVDAHYHEPBPBROSLLFOMENRSLCSAPACDILIPDEILIPRACHLDSQP
FEVEKACRSTSHVLEKVERFENLOSILYPLIKDPSLNHIVHLVLPDPAVERSRERT
GDLMDIRIVWGQHEOKLKKEDQSYVMQVLCQSLKTYKTIQSLPKLQERYLYLVRY
EDLARAVAOQTSRMVEFVLEFLPLQTMVHNITRGKMGMDHAFHTNARDLANTSOAM
RMSLPYKVSRLQACGDAMNLLGVRHVRSEQBNLLDLSTWYPEQIH"
CDS
BASE COUNT 451 a 542 c 474 g 512 t
ORIGIN
Query Match 75.7%; Score 1547; DB 9; Length 1979;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 106 CAAGGCTTCACCTTCAGCAACAATGCTACCTGCTTAAAAAATGAAGCTCGCTGCTTCT 165
Db 47 CAAGGCTTCACCTTCAGCAACAATGCTACCTGCTTAAAAAATGAAGCTCGCTGCTTCT 106
QY 166 GGTTCACAGATGAGCAATCTGGCTCTATCTTCCACATGTACAGCCACAACATCAGCTC 225
Db 107 GGTTCACAGATGAGCAATCTGGCTCTATCTTCCACATGTACAGCCACAACATCAGCTC 166
QY 226 CCTGTCTATGAGCACAAGCCGAGCGCATGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 285
Db 167 CCTGTCTATGAGCACAAGCCGAGCGCATGACGCTGCTGCTGCTGCTGCTGCTGCTGCT 226
QY 286 TGGCTCTCTTTTGGGGGAGCTTTTGGGACAGACCAACCAATGTTTCTACGATGAGA 345
Db 227 TGGCTCTCTTTTGGGGGAGCTTTTGGGACAGACCAACCAATGTTTCTACGATGAGA 286

QY 346 GCCCGCTGGACAGTGTGATGACCTTCAAGACAGACACCGCGTGTATGTCACATGGC 405
Db 287 GCCCGCTGGACAGTGTGATGACCTTCAAGACAGACACCGCGTGTATGTCACATGGC 346
QY 406 TGTGGGATGATGATAGGGGCGCTCTTGTGACATGAGACGCTTGTGATGCTACAT 465
Db 347 TGTGGGATGATGATAGGGGCGCTCTTGTGACATGAGACGCTTGTGATGCTACAT 406
QY 466 GGAACCTGTGTCGGGAGACAGTCCAGGCTTTCATAGTGGAGAACACCGGCGCTGTG 525
Db 407 GGAACCTGTGTCGGGAGACAGTCCAGGCTTTCATAGTGGAGAACACCGGCGCTGTG 466
QY 526 TTCTGACCTGCTGTGACATCATCCCAAGATGAATATCCCGGGGCTGACATGACAG 585
Db 467 TTCTGACCTGCTGTGACATCATCCCAAGATGAATATCCCGGGGCTGACATGACAG 526
QY 586 GCTCCTGTGACATCAACAGCCCTTGTAGGTGTGAGAGAGGCGCTGCTCTACAGCA 645
Db 527 GCTCCTGTGACATCAACAGCCCTTGTAGGTGTGAGAGAGGCGCTGCTCTACAGCA 586
QY 646 CGTGTGCTCAAGAGGTGCGCTTCTTAACCTGACATCCCTACCGGCTGCTAAGA 705
Db 587 CGTGTGCTCAAGAGGTGCGCTTCTTAACCTGACATCCCTACCGGCTGCTAAGA 646
QY 706 CCCCCTCCCAACCTGACATTCGTGACCTGCTGCGGAGCCCGGGGCGGTTCGCTTC 765
Db 647 CCCCCTCCCAACCTGACATTCGTGACCTGCTGCGGAGCCCGGGGCGGTTCGCTTC 706
QY 766 CCGAGACGACAAAGGAGATCTCATGATGACAGTGTGATGATGGGACACATGA 825
Db 707 CCGAGACGACAAAGGAGATCTCATGATGACAGTGTGATGATGGGACACATGA 766
QY 826 GCAAAACTCAAGAGAGAGACCAACCTTACTATGTATGACAGTCTATCTGCCAAAGCA 885
Db 767 GCAAAACTCAAGAGAGAGACCAACCTTACTATGTATGACAGTCTATCTGCCAAAGCA 826
QY 886 GCTGAGATCTACAGACATCATCAGTCTGCGCAAGGCGCGGAGAGAGCTACCTGCT 945
Db 827 GCTGAGATCTACAGACATCATCAGTCTGCGCAAGGCGCGGAGAGAGCTACCTGCT 886
QY 946 TGTGGCTATGAGACCTGTGCTGAGCCCTGTGCGCCAGACTTCCGAAATGTATGAAT 1005
Db 887 TGTGGCTATGAGACCTGTGCTGAGCCCTGTGCGCCAGACTTCCGAAATGTATGAAT 946
QY 1006 CTTGGGATTTGGAATCTTTCGCCATCTTCAAGACCTGCGTGCATTAACATCACCGGCA 1065
Db 947 CTTGGGATTTGGAATCTTTCGCCATCTTCAAGACCTGCGTGCATTAACATCACCGGCA 1006
QY 1066 GGGCATGGGTGACACAGCTTTCACACAATAATCCAGGAGTGGCCTTAATGTCTCCAGGC 1125
Db 1007 GGGCATGGGTGACACAGCTTTCACACAATAATCCAGGAGTGGCCTTAATGTCTCCAGGC 1066
QY 1126 TTGGCGTGTGCTTTCCTATGAAAAGTTCCTGACTTGAGAAAGCTGTGGGATGC 1185
Db 1067 TTGGCGTGTGCTTTCCTATGAAAAGTTCCTGACTTGAGAAAGCTGTGGGATGC 1126
QY 1186 CATGAATTTGCTGGGCTACCGGACAGTCAAGATCTTAACAAGAAAGAGAAACCTTGTCT 1245
Db 1127 CATGAATTTGCTGGGCTACCGGACAGTCAAGATCTTAACAAGAAAGAGAAACCTTGTCT 1186
QY 1246 GGATCTTGTCTACCTGAGCTGTCCCTGAGACAATTCACATGAAGGGTTGAGAGGCTT 1305
Db 1187 GGATCTTGTCTACCTGAGCTGTCCCTGAGACAATTCACATGAAGGGTTGAGAGGCTT 1246
QY 1306 TGTGACCACTGAGTGTACAGCTCACTCACTTCTCTGAATGCTTGAAGCTTGCATCA 1365
Db 1247 TGTGACCACTGAGTGTACAGCTCACTCACTTCTCTGAATGCTTGAAGCTTGCATCA 1306
QY 1366 TCTGAGACCTTAACATCATGCTGTGGGTATCAACATGATGATGATGATGATGATGATG 1425
Db 1307 TCTGAGACCTTAACATCATGCTGTGGGTATCAACATGATGATGATGATGATGATGATG 1366

QY 1426 TGTCTAAGCAGAGAGACTTTTGTCTCAGTCTGTCTAGAAAACAGACTGGGGACCT 1485
DB 1367 TGTCTAAGCAGAGAGACTTTTGTCTCAGTCTGTCTAGAAAACAGACTGGGGACCT 1426
QY 1486 TATGTAGAGAGACATCCACGAGTGAACAGGGTATGTCTCTCTCTTTCTTCTATCT 1545
DB 1427 TATGTAGAGAGACATCCACGAGTGAACAGGGTATGTCTCTCTCTTTCTTCTATCT 1486
QY 1546 TCTGTCTGTGGGAGAGACTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1605
DB 1487 TCTGTCTGTGGGAGAGACTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1546
QY 1606 ATCAGTGAATGATATCATTAACCTCCCTGTCTCAGATCTTGTCCCAATGGGAATGATCT 1665
DB 1547 ATCAGTGAATGATATCATTAACCTCCCTGTCTCAGATCTTGTCCCAATGGGAATGATCT 1606
QY 1666 TTCACCAAGAGCTCACAGACTTTTCCACAGAGATGC 1703
DB 1607 TTCACCAAGAGCTCACAGACTTTTCCACAGAGATGC 1644

RESULT 8
AR203335 AR203335 2032 bp DNA linear PAT 20-JUN-2002
LOCUS Sequence 2 from patent US 6365365.
DEFINITION AR203335
ACCESSION AR203335
VERSION AR203335.1 GI:21496698
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistrup, A., Rosen, S.D., Tangemann, K. and Hemmerich, S.
TITLE Method of determining whether an agent modulates glycosyl
sulfotransferase-3
JOURNAL Patent: US 6365365-A 2 02-Apr-2002;
FEATURES
source Location/Qualifiers
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN

Query Match 75.7%; Score 1547; DB 6; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 526 TTCTGACCTGCTGTGATCATCCACAGATGAATCATCCCGGGCTCAGTGCAG 585
DB 580 TTCTGACCTGCTGTGATCATCCACAGATGAATCATCCCGGGCTCAGTGCAG 639
QY 586 GCTCTGTGTGATCAACAGCCCTTTGAGTGTGTGAGAGAGAGAGAGAGAGAGAGAG 645
DB 640 GCTCTGTGTGATCAACAGCCCTTTGAGTGTGTGAGAGAGAGAGAGAGAGAGAGAG 699
QY 646 GCTGTGTGTGAG 705
DB 700 GCTGTGTGTGAG 759
QY 706 CCCCTCCCTCAGCTGATATCGTGCACCTGATCCGGAGACCCCGGGCCGTCTCCGTTTC 765
DB 760 CCCCTCCCTCAGCTGATATCGTGCACCTGATCCGGAGACCCCGGGCCGTCTCCGTTTC 819
QY 766 CCGAGAAAGCAGAAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 825
DB 820 CCGAGAAAGCAGAAAGAGAGATGATGATGATGATGATGATGATGATGATGATGATGAT 879
QY 826 GCAAAAACCTCAAG 885
DB 880 GCAAAAACCTCAAG 939
QY 886 GCTGTGATCTACAGAGACATCCAGTCTTGTCCCAAGAGAGAGAGAGAGAGAGAGAG 945
DB 940 GCTGTGATCTACAGAGACATCCAGTCTTGTCCCAAGAGAGAGAGAGAGAGAGAGAG 999
QY 946 TGT 1005
DB 1000 TGT 1059
QY 1006 CGT 1065
DB 1060 CGT 1119
QY 1066 GGGCATGGGTGACACAGCTTTCCACACAAATGCGAGAGATGCCCTTAATGTCTCCAGGC 1125
DB 1120 GGGCATGGGTGACACAGCTTTCCACACAAATGCGAGAGATGCCCTTAATGTCTCCAGGC 1179
QY 1126 TTGGCGCTGT 1185
DB 1180 TTGGCGCTGT 1239
QY 1186 CATGAATTTGCTGT 1245
DB 1240 CATGAATTTGCTGT 1299
QY 1246 GATCTTTGT 1305
DB 1300 GATCTTTGT 1359
QY 1306 TGT 1365
DB 1360 TGT 1419
QY 1366 TCTGTAGAGCTTACTAGATGT 1425
DB 1420 TCTGTAGAGCTTACTAGATGT 1479
QY 1426 TGTCTAAGCAG 1485
DB 1480 TGTCTAAGCAG 1539
QY 1486 TATGTAG 1545
DB 1540 TATGTAG 1599
QY 1546 TCTGTCTGTGGGAGAGACTTGTGAG 1605
DB 1600 TCTGTCTGTGGGAGAGACTTGTGAG 1659

QY 1606 ATCAGTGAATGATTCATCAACCTCCCTGTCACATCTTGCCCAATGGGGAATGATCT 1665
|||||
Db 1660 ATCAGTGAATGATTCATCAACCTCCCTGTCACATCTTGCCCAATGGGGAATGATCT 1719
|||||
QY 1666 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1703
|||||
Db 1720 TTCACCAAGAGCTCACCAGCATTTTCCACAGAGATGC 1757
|||||
RESULT 9
BD134772 2032 bp DNA linear PAT 18-SEP-2002
LOCUS Glycosylsulfotransferase-3.
DEFINITION BD134772
ACCESSION BD134772.1 GI:23229717
VERSION BD134772.1
KEYWORDS JP 2002507409-A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homiidae; Homo.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistrup, A., Rosen, S.D., Tangemann, K. and Hemmerich, S.
TITLE Glycosylsulfotransferase-3
JOURNAL Patent: JP 2002507409-A 1 12-MAR-2002;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, SYNTAX INC
OS Homo sapiens (human)
PN JP 2002507409-A/1
PD 12-MAR-2002
PR 26-FEB-1999 JP 2000537979
PR 20-MAR-1998 US 09/045284, 12-NOV-1998 US 09/190911 PI
ANETTE BISTRUP, STEVEN D ROSEN, KIRSTEN TANGEMANN, STEFAN PI
HEMMERICH
PC C12N9/10,A01K67/027,A61K45/00,A61K48/00,A61P29/00,A61P37/06,
PC A61P43/00,
PC C12N1/15,C12N1/21,C12N5/10,C12N15/09,C12O1/48,C12N5/00,C12N15/
PC 00
CC Glycosylsulfotransferase-3
FH Key Location/Qualifiers
FT source 1..2032
FT location/Qualifiers
1..2032
1..2032
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN
Query Match 75.7%: Score 1547; DB 6; Length 2032;
Best Local Similarity 99.9%: Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 106 CAAGGTCTCCACTGACGACATGCTACGCGCCAAAATGAAGTCTCTGCTTCT 165
|||||
Db 160 CAAGGTCTCCACTGACGACATGCTACGCGCCAAAATGAAGTCTCTGCTTCT 219
|||||
QY 166 GGTTCACAGATGGCATCTTGCTCTATCTTCCACATGTACAGCACAACATCAGCTC 225
|||||
Db 220 GGTTCACAGATGGCATCTTGCTCTATCTTCCACATGTACAGCACAACATCAGCTC 279
|||||
QY 226 CCGTCTATGAAGCAGCAGCCGAGCGCATGACGTCGTCTGCTTCTGCGCCTC 285
|||||
Db 280 CCGTCTATGAAGCAGCAGCCGAGCGCATGACGTCGTCTGCTTCTGCGCCTC 339
|||||
QY 286 TGCTCTCTTTTGTGGGGAGCTTTTGGGAGCAGCCAGATCTTTTACCTGATGA 345
|||||
Db 340 TGCTCTCTTTTGTGGGGAGCTTTTGGGAGCAGCCAGATCTTTTACCTGATGA 399
|||||
QY 346 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGCAGCCGCTGATGCTGACATGGC 405
|||||
Db 400 GCCCGCTGGCAGCTGTGATGACCTTCAAGCAGAGCAGCCGCTGATGCTGACATGGC 459
|||||
QY 406 TGTGGGATCTGATACGGGCGCTTCTTGTGGACATGAGCGCTTTGATGCTACAT 465
|||||

Db 460 TGTGGGATCTGATACGGGCGCTTCTTGTGGACATGAGCGCTTTGATGCTACAT 519
|||||
QY 466 GGAACCTGTGCCCGGAGACAGTCCAGCCTTTTCACTGGAGAACAGCGGCGCTGTG 525
|||||
Db 520 GGAACCTGTGCCCGGAGAGAGTCCAGCCTTTTCACTGGAGAACAGCGGCGCTGTG 579
|||||
QY 526 TTTGCAACCTGGCGGTGACATCATCCCAAGATGAATATCCCGGGGCTCACTGAG 585
|||||
Db 580 TTTGCAACCTGGCGGTGACATCATCCCAAGATGAATATCCCGGGGCTCACTGAG 639
|||||
QY 586 GCTCTGTGAGTCAACAGCCCTTTGAGTGTGGAGAGAGCCCTCCCTCTACAGCA 645
|||||
Db 640 GCTCTGTGAGTCAACAGCCCTTTGAGTGTGGAGAGAGCCCTCCCTCTACAGCA 699
|||||
QY 646 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTAAGA 705
|||||
Db 700 CGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTAAGA 759
|||||
QY 706 CCCCTCCCTCAACCTGATATCGTACACCTGTGCGGAGACCCCGGCGCTTCCGTTG 765
|||||
Db 760 CCCCTCCCTCAACCTGATATCGTACACCTGTGCGGAGACCCCGGCGCTTCCGTTG 819
|||||
QY 766 CCGAAGACGACAAAGGAGATCTCATGATTGACAGTCCGATTGTGGGAGACATGA 825
|||||
Db 820 CCGAAGACGACAAAGGAGATCTCATGATTGACAGTCCGATTGTGGGAGACATGA 879
|||||
QY 826 GCAAAACTCAAGAGAGAGACCAACCCCTACTATGTATGATGAGTCACTGCCAAAGCA 885
|||||
Db 880 GCAAAACTCAAGAGAGAGACCAACCCCTACTATGTATGATGAGTCACTGCCAAAGCA 939
|||||
QY 886 GCTGAGATCTACAAAGCATATCCAGTCTTGCCTCAAGGAGCCCTGAGAGACGCTACTGCT 945
|||||
Db 940 GCTGAGATCTACAAAGCATATCCAGTCTTGCCTCAAGGAGCCCTGAGAGACGCTACTGCT 999
|||||
QY 946 TGTGCGTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCGAATGTGATTT 1005
|||||
Db 1000 TGTGCGTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCGAATGTGATTT 1059
|||||
QY 1006 CGTGGGATGGAATCTTGGCCCATCTTGACAGCTGGGTGATTAACACCCGAGGCA 1065
|||||
Db 1060 CGTGGGATGGAATCTTGGCCCATCTTGACAGCTGGGTGATTAACACCCGAGGCA 1119
|||||
QY 1066 GGGCATGGTACACACCTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGC 1125
|||||
Db 1120 GGGCATGGTACACACCTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCAGGC 1179
|||||
QY 1126 TTTGGCGTGTCTTGGCCTATGAAAGTTTCTGACCTTCAGAAAGCTGTGCGATGC 1185
|||||
Db 1180 TTTGGCGTGTCTTGGCCTATGAAAGTTTCTGACCTTCAGAAAGCTGTGCGATGC 1239
|||||
QY 1186 CATGAATTTGCTGGGCTACCGCCAGCTGATGATGAAACAGAGAAACCGTTTGTCT 1245
|||||
Db 1240 CATGAATTTGCTGGGCTACCGCCAGCTGATGATGAAACAGAGAAACCGTTTGTCT 1299
|||||
QY 1246 GGATCTTCTGTCTACTGAGTGTCTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTT 1305
|||||
Db 1300 GGATCTTCTGTCTACTGAGTGTCTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTT 1359
|||||
QY 1306 TGTGCAACCTGTGTAGGCTCAGTCACTTCTGTGAACTTCTGAGCTTGCCTTACA 1365
|||||
Db 1360 TGTGCAACCTGTGTAGGCTCAGTCACTTCTGTGAACTTCTGAGCTTGCCTTACA 1419
|||||
QY 1366 TGTGCAACCTTACTCATGATGTGTGTGGGTATCATCATGATGATGTGTGCTACAG 1425
|||||
Db 1420 TGTGCAACCTTACTCATGATGTGTGTGGGTATCATCATGATGATGTGTGCTACAG 1479
|||||
QY 1426 TGTCAAGCAGAAAGACTTTTGTGTCCATGCTTGTCTGTAAGAAACAGACTGGGAACT 1485
|||||
Db 1480 TGTCAAGCAGAAAGACTTTTGTGTCCATGCTTGTGTCTGTAAGAAACAGACTGGGAACT 1539
|||||
QY 1486 TATGTGACAGACATCCCAACCAAGGATGATGCTTCTTCTTGTGATCT 1545
|||||

Db 1540 TATGAGACGACATCCACACAGGGAATTCCTCTCTTTCTTGATCT 1599
QY 1546 TCCTGTCTGGGAGAGACTTGTGGCTGGAGCCATTAAGCAGACAGT 1605
Db 1600 TCCTGTCTGGGAGAGACTTGTGGCTGGAGCCATTAAGCAGACAGT 1659
QY 1606 ATCAGTGGAAATGATTCATTAACCTCCCTGTGCACATCTTGCCCAATGGGAATGATCT 1665
Db 1660 ATCAGTGGAAATGATTCATTAACCTCCCTGTGCACATCTTGCCCAATGGGAATGATCT 1719
QY 1666 TTCACCAAGAGCTCACAGCATTTTCCACAGATGC 1703
Db 1720 TTCACCAAGAGCTCACAGCATTTTCCACAGATGC 1757

RESULT 10
AF131235 2032 bp mRNA linear PRI 10-AUG-1999
LOCUS AF131235
DEFINITION Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA,
complete cds.
ACCESSION AF131235
VERSION AF131235.1 GI:4927113
KEYWORDS
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistrup, A., Bhakta, S., Lee, J.-K., Belov, Y. Y., Gunn, M. D., Zuo, F. R.,
Huang, C. C., Kannagi, R., Rosen, S. D. and Hemmerich, S.
Sulfotransferases of two specificities function in the
reconstitution of high endothelial cell ligands for L-selectin
J. Cell Biol. 145 (4), 899-910 (1999)
JOURNAL 99264336
MEDLINE 10330415
PUBMED 2 (bases 1 to 2032)
REFERENCE Bistrup, A., Tangemann, K., Bhakta, S., Lee, J.-K., Belov, Y. Y.,
Gunn, M. D., Zuo, F. R., Huang, C. C., Kannagi, R., Rosen, S. D. and
Hemmerich, S.
Direct Submision
Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience,
3401 Hillview Avenue, Palo Alto, CA 94304, USA
FEATURES
Source location/Qualifiers
1..2032
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/cell_type="high endothelial"
/tissue_type="tonsil"
182..1342
/function="biosynthesis of endothelial ligands for
L-selectin"
/note="HEC-G1CNAC-6-SY"
/product="N-acetylglucosamine 6-O-sulfotransferase"
/protein_id="AAD33015.1"
/db_xref="GI:4927113"
translation="MLPRLKMLLFLVSQMLALFFMHYSHNISLSMKAKPERH
VLVLSWRGSEFVQLFGQHPDVLYLMPHAWMTTEFCSTAMMIHNAVRLIAVE
LIDMSVPDAEMEPGRROSLFOMENSRLCSAPADIPODEIIPRAICRLISQOP
FEVFKACRSYSHVYLKEVREFNLSIVPLKDPNINILVLDPRAPVRSREPTK
GLMIDSRIVMGHOKLKEKEDQPYVMAVYIQSOLSEITKIQSLPKALQERILVRY
EDLAPAVAQTSRMTFVGLFELPHLQIVWHNITKSGKGDIAFTNADALNVSOAW
RMSLPEVYSRLQKACGDAMNLLGYRHVSEOEORNLIDLSTLTVTPOIH"

BASE COUNT 468 a 569 c 490 g 505 t

ORIGIN
Query Match 75.7%; Score 1547; DB 9; Length 2032;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1597; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 106 CAAGGCTTCACCTCAGACAAATGCTACTGCTAAATAAATGAAGCTCCTGCTGTTCT 165
|||||

Db 160 CAAGGCTTCACCTCAGACAAATGCTACTGCTTAAAAAATGAAGCTCCTGCTGTTCT 219
QY 166 GGTTCACAGATGGGACATCTTGAGCTATCTCTCCACATGATACAGGACACAAATCAGCTC 225
Db 220 GGTTCACAGATGGGACATCTTGAGCTATCTCTCCACATGATACAGGACACAAATCAGCTC 279
QY 226 CCGTGTATGAAAGGACACAGCCGAGCGATGACAGCTGTGTGTCTCTTCTGCGCTC 285
Db 280 CCGTGTATGAAAGGACACAGCCGAGCGATGACAGCTGTGTGTCTCTTCTGCGCTC 339
QY 286 TGGCTCTCTTTTGTGGGAGAGCTTTTGGGAGAGCCAGATGTTTCTACTGATGGA 345
Db 340 TGGCTCTCTTTTGTGGGAGAGCTTTTGGGAGAGCCAGATGTTTCTACTGATGGA 399
QY 346 GCCCGCTGGACAGTGTGATGATCCCTTAAGCAGACAGCCGCTGATGTCACATGAC 405
Db 400 GCCCGCTGGACAGTGTGATGATCCCTTAAGCAGACAGCCGCTGATGTCACATGAC 459
QY 406 TGTGGGGATCTGATACGGGCGCTCTTGTGTGACATGAGCGTCTTTGATGCTTACAT 465
Db 460 TGTGGGGATCTGATACGGGCGCTCTTGTGTGACATGAGCGTCTTTGATGCTTACAT 519
QY 466 GGAACCTGGTCCCGGAGAGAGTCCAGCCCTTTTCACTGGGAGAAAGCCGGGCGCTGTG 525
Db 520 GGAACCTGGTCCCGGAGAGAGTCCAGCCCTTTTCACTGGGAGAAAGCCGGGCGCTGTG 579
QY 526 TTCTGCACCTGCTGTGACATCATCCACAAAGATGAATCATCCCGGGGCTCCTACGAG 585
Db 580 TTCTGCACCTGCTGTGACATCATCCACAAAGATGAATCATCCCGGGGCTCCTACGAG 639
QY 586 GCTCTGTGACGTAACAGCCCTTTGAGGTGTGAGAAAGCCCTGCGCTCTTACAGCA 645
Db 640 GCTCTGTGACGTAACAGCCCTTTGAGGTGTGAGAAAGCCCTGCGCTCTTACAGCA 699
QY 646 CGTGGTCAAGGAGAGTGGGCTTTCAACCTGAGTCCCTTACCGCTGTGAAAGA 705
Db 700 CGTGGTCAAGGAGAGTGGGCTTTCAACCTGAGTCCCTTACCGCTGTGAAAGA 759
QY 706 CCCCTCCCTCAACCTGATCGTGCACCTGTGTCGGGAGCCCGGGCGGTGTCCGTT 765
Db 760 CCCCTCCCTCAACCTGATCGTGCACCTGTGTCGGGAGCCCGGGCGGTGTCCGTT 819
QY 766 CCGAAGACGACAAAGGAGATCTCATGATGACATCGCATTTGTATGGGAGCATGA 825
Db 820 CCGAAGACGACAAAGGAGATCTCATGATGACATCGCATTTGTATGGGAGCATGA 879
QY 826 GCAAAAACGCAAGAGAGAGACCAACCTACTATGATGATGACAGTATGTCGCAAGCA 885
Db 880 GCAAAAACGCAAGAGAGAGACCAACCTACTATGATGATGACAGTATGTCGCAAGCA 939
QY 886 GCTGAGATCTACAGACATCCAGTCTTGCCCAAGGCCCTGACAGAAAGCTACCTGCT 945
Db 940 GCTGAGATCTACAGACATCCAGTCTTGCCCAAGGCCCTGACAGAAAGCTACCTGCT 999
QY 946 TGTGGCTATGAGAGACCTGGCTGAGCCCTGTGGGCCAGACTCCGGAATGATGAAT 1005
Db 1000 TGTGGCTATGAGAGACCTGGCTGAGCCCTGTGGGCCAGACTCCGGAATGATGAAT 1059
QY 1006 CCGTGGATGGAATCTTCCCATCTCAGACCTGGGTGATATCATCATCCGAGGAA 1065
Db 1060 CCGTGGATGGAATCTTCCCATCTCAGACCTGGGTGATATCATCATCCGAGGAA 1119
QY 1066 GGGCATGGGTGACACAGCTTCCACACAATGCGAGGATGCCCTTAATGTCTCCAGGC 1125
Db 1120 GGGCATGGGTGACACAGCTTCCACACAATGCGAGGATGCCCTTAATGTCTCCAGGC 1179
QY 1126 TTGGCGCTGTCTTGGCCCTATGAAGAAGTTTCTGACTTCGAAGAAGCTGTGGCATGC 1185
Db 1180 TTGGCGCTGTCTTGGCCCTATGAAGAAGTTTCTGACTTCGAAGAAGCTGTGGCATGC 1239
QY 1186 CATGAATTTGCTGGGTACCGCACGCTCATGATCTGAACAAGAGAGAACTGTGTCT 1245
Db 1240 CATGAATTTGCTGGGTACCGCACGCTCATGATCTGAACAAGAGAAACTGTGTCT 1299

QY 1246 GGATCTCTCTACTGAGCTGCTGAGCAATCCAGTAGAGGCTT 1305
 |||||
 Db 1300 GGATCTCTCTCTACTGAGCTGCTGAGCAATCCAGTAGAGGCTT 1359
 |||||
 QY 1306 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1365
 |||||
 Db 1360 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1419
 |||||
 QY 1366 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1425
 |||||
 Db 1420 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1479
 |||||
 QY 1426 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1485
 |||||
 Db 1480 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1539
 |||||
 QY 1486 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1545
 |||||
 Db 1540 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1599
 |||||
 QY 1546 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1605
 |||||
 Db 1600 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1659
 |||||
 QY 1606 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1665
 |||||
 Db 1660 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1719
 |||||
 QY 1666 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1703
 |||||
 Db 1720 TGGTCCACCTGCTGAGCTGAGCTGAGCTGAGCTGAGCTGAGCT 1757
 |||||

RESULT 11
 AF149783 1333 bp mRNA linear PRI 02-JUL-2001
 LOCUS Homo sapiens L-selectin ligand sulfotransferase mRNA, complete cds.
 AF149783
 AF149783.1 GI:13897503
 VERSION
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.
 (bases 1 to 1333)
 Yeh,J.C., Hiraoaka,N., Petryniak,B., Nakayama,J., Ellices,L.G.,
 Rebuka,D., Hindsgraul,O., Marth,J.D., Lowe,J.B. and Fukuda,M.,
 Novel sulfated lymphocyte homing receptors and their control by a
 Core1 extension beta 1,3-N-acetylglucosaminyltransferase
 Cell 105 (7), 957-969 (2001)

REFERENCE
 AUTHORS
 TITLE
 JOURNAL
 MEDLINE
 PUBMED
 21332592
 11439191
 2 (bases 1 to 1333)
 Hiraoaka,N. and Fukuda,M.
 Direct Submission
 Submitted (10-MAY-1999) Gynecobiology, The Burnham Institute, 10901
 North Torrey Pines Road, La Jolla, CA 92037, USA
 location/qualifiers
 1. 1333
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 93. 1253
 /note="LST"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase"
 /protein_id="AAK48417.1"
 /db_xref="GI:13897504"
 /translation="MLPKRMKLLFLVSMATLALFPHYSHNLSLSKAOERH
 VLIVSRSSSSPYGOLFQGHDPVFLMEPRMHWMTFKSTAMLMAYADLRAVY
 LQDMSVDATMEPRKSSSLFQWENRSLACSAPACDITLQDEITIPAHKRLCSQD
 FEVEKACRSYSHVVLKVEVFENFLOSILYPLKPSLMLIHVLVDRPRAVRSRERK

CDS

1. 1333
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 93. 1253
 /note="LST"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase"
 /protein_id="AAK48417.1"
 /db_xref="GI:13897504"
 /translation="MLPKRMKLLFLVSMATLALFPHYSHNLSLSKAOERH
 VLIVSRSSSSPYGOLFQGHDPVFLMEPRMHWMTFKSTAMLMAYADLRAVY
 LQDMSVDATMEPRKSSSLFQWENRSLACSAPACDITLQDEITIPAHKRLCSQD
 FEVEKACRSYSHVVLKVEVFENFLOSILYPLKPSLMLIHVLVDRPRAVRSRERK

BASE COUNT 286 a 393 c 336 g 318 t
 ORIGIN
 Query Match 61.8%; Score 1262; DB 9; Length 1333;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 GMLIDSRIVMGOEOKLKKEDDPYVVMQVICOSEIYKTIQSLPRALOERYLVRY
 EDLAPRAVQTSRYAEVGEFLPHQIOTVHTNRGMDHAFHTNARALNVSOAM
 RMLSPYKVRSLQKAGSDAMNLLGRHVRSEQRNLLDLLSTWVPEQIH"
 107 AAGGTCTCCACTTCACACACATGCTACTGCTAAAAAATGAAGCTCTGCTTTCTG 166
 |||||
 Db 72 AAGGTCTCCACTTCACACACATGCTACTGCTAAAAAATGAAGCTCTGCTTTCTG 131
 |||||
 QY 167 GTTCCAGATGGCCATCTGGCTCTATGCTCCACATGTACAGCACAACATAGCTCC 226
 |||||
 Db 132 GTTCCAGATGGCCATCTGGCTCTATGCTCCACATGTACAGCACAACATAGCTCC 191
 |||||
 QY 227 CTGTCTATGAAGGACAGCCGAGCGATGACAGCTGCTGCTTCTGCTGCGCTCT 286
 |||||
 Db 192 CTGTCTATGAAGGACAGCCGAGCGATGACAGCTGCTGCTTCTGCTGCGCTCT 251
 |||||
 QY 287 GGTCTCTCTTTTGGGGGACCTTTTGGGAGCAGCAGCAGATTTTCTACTATAGAG 346
 |||||
 Db 252 GGTCTCTCTTTTGGGGGACCTTTTGGGAGCAGCAGCAGATTTTCTACTATAGAG 311
 |||||
 QY 347 CCGGCTGGGACAGCTGGATGACCTTCAGAGAGCAGCCGCTGGATGCTGACATGAGCT 406
 |||||
 Db 312 CCGGCTGGGACAGCTGGATGACCTTCAGAGAGCAGCCGCTGGATGCTGACATGAGCT 371
 |||||
 QY 407 GTGGGGATCTGATACGGGCGCTTCTGTGCGACATGACGCTTTGATGCTTACATG 466
 |||||
 Db 372 GTGGGGATCTGATACGGGCGCTTCTGTGCGACATGACGCTTTGATGCTTACATG 431
 |||||
 QY 467 GAACCTGTCCTCCGAGACATCCAGCTCTTATAGTGGAGAGACCGGGCCCTGCT 526
 |||||
 Db 432 GAACCTGTCCTCCGAGACATCCAGCTCTTATAGTGGAGAGACCGGGCCCTGCT 491
 |||||
 QY 527 TCTGACCTGCTGATGATCATCCCAAGATGAATCATCCCGGGCTCACTGAGG 586
 |||||
 Db 492 TCTGACCTGCTGATGATCATCCCAAGATGAATCATCCCGGGCTCACTGAGG 551
 |||||
 QY 647 GTGTCGTCAGAGGAGGCGCTTCTCAACCTGAGAGTCCCTGACCGCTGTAAGAG 706
 |||||
 Db 612 GTGTCGTCAGAGGAGGCGCTTCTCAACCTGAGAGTCCCTGACCGCTGTAAGAG 671
 |||||
 QY 707 CCTCCCTCAACCTGATATGCTGACCTGCTCCGGGACCCCGGGCGCTTCCGTTCC 766
 |||||
 Db 672 CCTCCCTCAACCTGATATGCTGACCTGCTCCGGGACCCCGGGCGCTTCCGTTCC 731
 |||||
 QY 767 CGAGACGACAAAAGGAGATCTCATGATTGACAGTGGCATTTGATGAGGGGACAGTAG 826
 |||||
 Db 732 CGAGACGACAAAAGGAGATCTCATGATTGACAGTGGCATTTGATGAGGGGACAGTAG 791
 |||||
 QY 827 CAAAACCTCAAGAGAGAGACCAACCTTACTATGTGATGAGGATCTGCAAGAGCAG 886
 |||||
 Db 792 CAAAACCTCAAGAGAGAGACCAACCTTACTATGTGATGAGGATCTGCAAGAGCAG 851
 |||||
 QY 887 CTGAGATCTCAAGACATCATGATCTTGGCCCAAGGCCCTGCAAGAGCGTACTGCTT 946
 |||||
 Db 852 CTGAGATCTCAAGACATCATGATCTTGGCCCAAGGCCCTGCAAGAGCGTACTGCTT 911
 |||||
 QY 947 GTGGCTATGAGGACCTGGCTGAGCCCTGTGGCCAGACTTCCGAAATGATGAATTC 1006
 |||||
 Db 912 GTGGCTATGAGGACCTGGCTGAGCCCTGTGGCCAGACTTCCGAAATGATGAATTC 971
 |||||
 QY 1007 GTGGATTTGGAATCTTGGCCATCTTCAAGACCTGGGCTGCTAATCATCAGCGAGCAG 1066
 |||||
 Db 972 GTGGATTTGGAATCTTGGCCATCTTCAAGACCTGGGCTGCTAATCATCAGCGAGCAG 1031
 |||||

QY 1067 GGCATGGGTGACACAGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCT 1126
 |||||||
 DB 1032 GGCATGGGTGACACAGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCT 1091
 QY 1127 TGGCGCTGGTCTTTGCCCTATGAAAAGTTTCTCCAGCTTCAGAAACCTGTGGCATGCC 1186
 |||||||
 DB 1092 TGGCGCTGGTCTTTGCCCTATGAAAAGTTTCTCCAGCTTCAGAAACCTGTGGCATGCC 1151
 QY 1187 ATGAATTTCTGGGGCTACCGCCACGTCAGATCTGAACAAAGAACAGAAACCGTGTGCTG 1246
 |||||||
 DB 1152 ATGAATTTCTGGGGCTACCGCCACGTCAGATCTGAACAAAGAACAGAAACCGTGTGCTG 1211
 QY 1247 GATCTTCTGTACCTGACACTGTCCCTGAGCAATCCACTAAGAGGGTTGAGAAGGCTTT 1306
 |||||||
 DB 1212 GATCTTCTGTACCTGACACTGTCCCTGAGCAATCCACTAAGAGGGTTGAGAAGGCTTT 1271
 QY 1307 GCTGCCACCTGGTGTACAGCTTACATCATTCTTGAAATGCTTTCAGAGCTTGCCTACAT 1366
 |||||||
 DB 1272 GCTGCCACCTGGTGTACAGCTTACATCATTCTTGAAATGCTTTCAGAGCTTGCCTACAT 1331
 QY 1367 CT 1368
 ||
 DB 1332 CT 1333
 ||
 RESULT 12
 AR310485 2065 bp DNA linear PAT 12-JUN-2003
 LOCUS Sequence 26 from patent US 6558935.
 DEFINITION AR310485
 ACCESSION AR310485
 VERSION AR310485.1 GI:31703448
 KEYWORDS
 SOURCE Unknown.
 ORGANISM Unknown.
 REFERENCE 1 (bases 1 to 2065)
 AUTHORS Tang,Y.T., Corley,N.C., Guegler,K.J., Baughn,M.R., Lal,P., Yue,H.,
 Hillman,J.L. and Azimzai,Y.
 TITLE Human Transferrase proteins
 JOURNAL Patent: US 6558935-A 26 06-MAY-2003;
 FEATURES
 source 1. 2065
 location/Qualifiers
 BASE COUNT 466 a 575 c 491 g 533 t
 ORIGIN
 Query Match 56.4%; Score 1153; DB 6; Length 2065;
 Best local Similarity 99.6%; Pred. No. 0;
 Matches 1593; Conservative 0; Mismatches 4; Indels 2; Gaps 2;

DB 452 TGTGGGGATCTGATACGGGCGCTCTTCTGTGCGACATGAGCGTCTTGATGCCATAT 511
 QY 466 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTAGTGGAGAACAGCCGGGCGCTGTG 525
 |||||||
 DB 512 GGAACCTGGTCCCGGAGACAGTCCAGCCCTTTAGTGGAGAACAGCCGGGCGCTGTG 571
 QY 526 TTGTGCACCTGGCTGTGACATCATCCACAGATGAATCATCCCGGGGCTCATGCA 585
 |||||||
 DB 572 TTGTGCACCTGGCTGTGACATCATCCACAGATG -AATCATCCCGGGGCTCATGCA 630
 QY 586 GCTCCGTGTCAATCAACACCCCTTTG-AGTGGTGGAGAAAGGCTCCCTCTCAAGCC 644
 |||||||
 DB 631 GCTCCGTGTCAATCAACACCCCTTTGAAATTTGTGGAGAAACCTCCCTCTCAAGCC 690
 QY 645 AGTGTGTCTCAAGAGAGTGGCGCTTTTGAACCTGAGTCCCTCAACCTCCGCTGCTAAAG 704
 |||||||
 DB 691 AGTGTGTCTCAAGAGAGTGGCGCTTTTGAACCTGAGTCCCTCAACCTCCGCTGCTAAAG 750
 QY 705 ACCCTCCCTCAACCTGATATGTCACCTGGTCCGGAGACCCCGGGCGGTTCGTT 764
 |||||||
 DB 751 ACCCTCCCTCAACCTGATATGTCACCTGGTCCGGAGACCCCGGGCGGTTCGTT 810
 QY 765 CCGGAGAACGCAAAAGGAGATCCATGATGACAGTGGCAATTTGTATGGGGGACAGAG 824
 |||||||
 DB 811 CCGGAGAACGCAAAAGGAGATCCATGATGACAGTGGCAATTTGTATGGGGGACAGAG 870
 QY 825 ACAAATACTCAAGAAAGAGAGACCAACCTACTATGTATGAGAGTCAATCTGCAAAAGCC 884
 |||||||
 DB 871 ACGAGAACTCAAGAAAGAGAGACCAACCTACTATGTATGAGAGTCAATCTGCAAAAGCC 930
 QY 885 AGCTGAGATCTCAAGAAAGAGAGTCAATCTGCTGCGCAAGCCCTGCAAGAAAGCTTCTGC 944
 |||||||
 DB 931 AGCTGAGATCTCAAGAAAGAGAGTCAATCTGCTGCGCAAGCCCTGCAAGAAAGCTTCTGC 990
 QY 945 TTGTGCGATATGAGAGCTGGCTGCGAGCCCTGTGGGCCAGAGCTTCCGAATGATGAAT 1004
 |||||||
 DB 991 TTGTGCGATATGAGAGCTGGCTGCGAGCCCTGTGGGCCAGAGCTTCCGAATGATGAAT 1050
 QY 1005 TGTGGGATTTGAATTTTGGCCCATCTTCAAGACCTGGTGTGATTAACATCACCGAGCA 1064
 |||||||
 DB 1051 TGTGGGATTTGAATTTTGGCCCATCTTCAAGACCTGGTGTGATTAACATCACCGAGCA 1110
 QY 1065 AAGGCAATGGGTGACCAACGCTTTCCACACAAATGCCAGGATGCCCTTATGTCTCCAG 1124
 |||||||
 DB 1111 AAGGCAATGGGTGACCAACGCTTTCCACACAAATGCCAGGATGCCCTTATGTCTCCAG 1170
 QY 1125 CTGGCGCTGGTCTTGGCCATATGAAAAGTTTCTGCAATCTTGAAGAGCTGTGGCATG 1184
 |||||||
 DB 1171 CTGGCGCTGGTCTTGGCCATATGAAAAGTTTCTGCAATCTTGAAGAGCTGTGGCATG 1230
 QY 1185 CCATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAACAAAGAGAGAAACCTGTGTC 1244
 |||||||
 DB 1231 CCATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAACAAAGAGAGAAACCTGTGTC 1290
 QY 1245 TGGATCTTGTCTACCTGAGCTGCCGACCAATTCACAAAGGGTTAGAAAGGT 1304
 |||||||
 DB 1291 TGGATCTTGTCTACCTGAGCTGCCGACCAATTCACAAAGGGTTAGAAAGGT 1350
 QY 1305 TTGCTGCCACCTGGTGTGACAGCTCACTTCTTCTGCAATCTTCTGAGAGCTTGGCTTAC 1364
 |||||||
 DB 1351 TTGCTGCCACCTGGTGTGACAGCTCACTTCTTCTGCAATCTTCTGAGAGCTTGGCTTAC 1410
 QY 1365 ATCTGTAGCCCTTAACATACATGTGTGGGTATCACACTGAGTGTGATGTGTCCACAC 1424
 |||||||
 DB 1411 ATCTGTAGCCCTTAACATACATGTGTGGGTATCACACTGAGTGTGATGTGTCCACAC 1470
 QY 1425 GTGCTTAAGCAAGAGACTTTTGTGTCCATGCTTGTGTCTAAGAAAACACATCTGGGAAC 1484
 |||||||
 DB 1471 GTGCTTAAGCAAGAGACTTTTGTGTCCATGCTTGTGTCTAAGAAAACACATCTGGGAAC 1530
 QY 1485 TTATGTGACAGACATCCACACAGTGAAGAAAGGATTTGCTTCTTCTTCTTCTTGTATG 1544
 |||||||
 DB 1531 TTATGTGACAGACATCCACACAGTGAAGAAAGGATTTGCTTCTTCTTCTTGTATG 1590

QY 1545 TTCTCTGTGGGAGAGACTTGTGGCTGGAGCCCTATTAGACAGACAG 1604
 DB 1591 TTCTCTGTGGGAGAGACTTGTGGCTGGAGCCCTATTAGACAGACAG 1650
 QY 1605 TATCAGTGAATGATCCATPAACCTCCGTGCCATCTTGCCCAATGGGAATGATC 1664
 DB 1651 TATCAGTGAATGATCCATPAACCTCCGTGCCATCTTGCCCAATGGGAATGATC 1710
 QY 1665 TTTCACCAAGAGCTCCAGCATTTTCCACAGAGATGC 1703
 DB 1711 TTTCACCAAGAGCTCCAGCATTTTCCACAGAGATGC 1749
 RESULT 13
 LOCUS BD124832 877 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD124832
 VERSION BD124832.1 GI:23219777
 KEYWORDS JP 2002017375-A/263.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
 Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
 Koga,H.
 TITLE Primer for synthesizing full-length cDNA and use thereof
 JOURNAL Patent: JP 2002017375-A 263 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/263
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TERUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TEISUJI OTSUKI,HISASHI KOGA
 PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10,
 PC C12P21/02,C12O1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH key
 FT source Location/Qualifiers
 1. 877
 /organism='Homo sapiens (human)'.
 1. 877
 /location/Qualifiers
 1. 877
 /organism='Homo sapiens'
 /mol_type='genomic DNA'
 /db_xref='taxon:9606'
 BASE COUNT 183 a 266 c 221 g 203 t 4 others
 ORIGIN
 Query Match 34.0%; Score 695; DB 6; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 106 CAAGGCTTCACATTCAGACAAATGCTAAGAAAAATGAAGCTCGCTGTTCT 165
 DB 47 CAAGGCTTCACATTCAGACAAATGCTAAGAAAAATGAAGCTCGCTGTTCT 106
 QY 166 GGTTCACAGATGGCATCTTGCTATTTCTTCACATGATACAGCCACAATCAGCTC 225
 DB 107 GGTTCACAGATGGCATCTTGCTATTTCTTCACATGATACAGCCACAATCAGCTC 166
 QY 226 CCGTGTATGAAGCAGACGCCAGCCAGCCATGACAGCTGCTGTTCTCTGCGCCTC 285
 DB 167 CCGTGTATGAAGCAGACGCCAGCCAGCCATGACAGCTGCTGTTCTCTGCGCCTC 226

QY 286 TGCGTCTCTTTTGTGGGAGAGCTTTTGGGACACACCAGANTTTTCTACCTGATGA 345
 DB 227 TGCGTCTCTTTTGTGGGAGAGCTTTTGGGACACACCAGANTTTTCTACCTGATGA 286
 QY 346 GCCCGCTGGACAGCTGTGATGACCTTCAACAGACAGACCGCTGTGATCTGACATGCC 405
 DB 287 GCCCGCTGGACAGCTGTGATGACCTTCAACAGACAGACCGCTGTGATCTGACATGCC 346
 QY 406 TGTGGGAGATCTGATGATGAGGCGGCTCTTGTGGACATGATGAGCTTTGATGCTTCAT 465
 DB 347 TGTGGGAGATCTGATGATGAGGCGGCTCTTGTGGACATGATGAGCTTTGATGCTTCAT 406
 QY 466 GGAACCTGTGCTCCCGGAGACAGTCCAGACCTTCTTCACTGTGGAGAAACCGCGGCTGTG 525
 DB 407 GGAACCTGTGCTCCCGGAGACAGTCCAGACCTTCTTCACTGTGGAGAAACCGCGGCTGTG 466
 QY 526 TTTCGACACCTGCTGTGATCATCCCAAGATGAATATCCCGGCTGCTGCTGACAG 585
 DB 467 TTTCGACACCTGCTGTGATCATCCCAAGATGAATATCCCGGCTGCTGCTGACAG 526
 QY 586 GCTGCTGTGACATCAAGACCGCTTGTAGTGTGTGGAGAAAGCGCTGCTGCTGACAGCA 645
 DB 527 GCTGCTGTGACATCAAGACCGCTTGTAGTGTGTGGAGAAAGCGCTGCTGCTGACAGCA 586
 QY 646 CGTGTGCTCAAGAGAGGTGCGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTAAGA 705
 DB 587 CGTGTGCTCAAGAGAGGTGCGCTTCTTCAACCTGACAGTCCCTTACCCGCTGTAAGA 646
 QY 706 CCCCTGCTCAACCTGATATGATGACCTGTGCGGAGACCGCGGCTGCTGCTGCTG 765
 DB 647 CCCCTGCTCAACCTGATATGATGACCTGTGCGGAGACCGCGGCTGCTGCTGCTG 706
 QY 766 CCGAAGACGCAAGAGGAGATCTCATGATTTGA 800
 DB 707 CCGAAGACGCAAGAGGAGATCTCATGATTTGA 741
 RESULT 14
 LOCUS BD126950 877 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD126950
 VERSION BD126950.1 GI:23221895
 KEYWORDS JP 2002017375-A/2381.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
 Makamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
 Koga,H.
 TITLE Primer for synthesizing full-length cDNA and use thereof
 JOURNAL Patent: JP 2002017375-A 2381 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/2381
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TERUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TEISUJI OTSUKI,HISASHI KOGA
 PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10,
 PC C12P21/02,C12O1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH key
 FT source Location/Qualifiers
 1. 877
 /organism='Homo sapiens (human)'.
 1. 877
 /location/Qualifiers
 1. 877
 /organism='Homo sapiens'
 /mol_type='genomic DNA'
 /db_xref='taxon:9606'
 FEATURES

source 1..877
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

BASE COUNT 183 a 266 c 221 g 203 t 4 others

ORIGIN

Query Match 34.0%; Score 695; DB 6; Length 877;
 Best Local Similarity 100.0%; Pred. No. 0;
 Matches 695; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 106 CAGGCTTCACCTAGCACAATGCTACTGCTTAATAAAGAACTCTGCTGTTCT 165
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 47 CAAAGCTTCACCTAGCACAATGCTACTGCTTAATAAAGAACTCTGCTGTTCT 106
 166 GTTTTCCAGATAGGCACTTGGCTTATCTTCACATGATACAGCACAACATCAGCTC 225
 107 GGTTCCTCCAGATAGGCACTTGGCTTATCTTCACATGATACAGCACAACATCAGCTC 166
 QY 226 CCTGTCTATGAAGGACAGCCCGAGCGCATGCACTGCTGTTCTGCTGCGCTC 285
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 167 CCGTCTATGAAGGACAGCCCGAGCGCATGCACTGCTGTTCTGCTGCGCTC 226
 286 TGGCTCTCTTTTGTGTGGGAGCTTTTGTGGGAGCAGCAGATGTTTCTACTGATGA 345
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 227 TGGCTCTCTTTTGTGTGGGAGCTTTTGTGGGAGCAGCAGATGTTTCTACTGATGA 286
 QY 346 GCCCGCTGACAGCTGTGTGATGACCTTCAAGCAGAGCAGCAGCTGTGATGCTGACATGCG 405
 287 GCCCGCTGACAGCTGTGTGATGACCTTCAAGCAGAGCAGCAGCTGTGATGCTGACATGCG 346
 QY 406 TGTGGGGATCTGATACGGGGCGCTCTTGTGTGGGAGCAGCTGTGATGCTGACATG 465
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 347 TGTGGGGATCTGATACGGGGCGCTCTTGTGTGGGAGCAGCTGTGATGCTGACATG 406
 466 GGAACCTGTGTGCGGAGAGCTCAGCTCTTTCAGTGGGAGAACAGCGGCGCTGTG 525
 407 GGAACCTGTGTGCGGAGAGCTCAGCTCTTTCAGTGGGAGAACAGCGGCGCTGTG 466
 QY 526 TTCTGACCTGCTGTGACATCATCCACAGATGAATCATCCCGGCGCTGACATGACG 585
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 467 TTCTGACCTGCTGTGACATCATCCACAGATGAATCATCCCGGCGCTGACATGACG 526
 586 GGTCCGTGACATCAACAGCCCTTGTGTGGGAGGAGCGCTGCGCTGCTACAGCACA 645
 527 GGTCCGTGACATCAACAGCCCTTGTGTGGGAGGAGCGCTGCGCTGCTACAGCACA 586
 QY 646 CGTGTGCTCAAGGAGGTGCGCTTCTCAACTGACAGTCCCTTACCCCTGCTGAAGA 705
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 587 CGTGTGCTCAAGGAGGTGCGCTTCTCAACTGACAGTCCCTTACCCCTGCTGAAGA 646
 706 CCCCTGCTCAACCTGACATATGCTGACCTGCTGCGGAGCCCGGCGCTGCTGCTG 765
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 647 CCCCTGCTCAACCTGACATATGCTGACCTGCTGCGGAGCCCGGCGCTGCTGCTG 706
 QY 766 CCGAGAGCGACAAGGAGATCTCATGATTGACA 800
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 707 CCGAGAGCGACAAGGAGATCTCATGATTGACA 741

RESULT 15
 AX381256 517 bp DNA 1linear PAT 18-MAR-2002
 LOCUS AX381256
 DEFINITION Sequence 194 from Patent WO0212280.
 ACCESSION AX381256
 VERSION AX381256.1 GI:19576075
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE
 1 Pyle, R.A., Xu, J. and Secrist, H.

TITLE Compositions and methods for the therapy and diagnosis of colon cancer
 JOURNAL Patent: WO 0212280-A 194 14-FEB-2002;
 CORIAX CORPORATION (US)
 FEATURES
 source 1..517
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"

BASE COUNT 121 a 130 c 124 g 142 t

ORIGIN

Query Match 19.7%; Score 403; DB 6; Length 517;
 Best Local Similarity 100.0%; Pred. No. 3.5e-228;
 Matches 403; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1301 GGCTTGCGCCACCGCTGGGTAGCCTCAGTCACTTTCCTAATGCTTCTGAGCCTTGC 1360
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 13 GGCTTGCGCCACCGCTGGGTAGCCTCAGTCACTTTCCTAATGCTTCTGAGCCTTGC 72
 QY 1361 CTACATCTGAGCCTTAATCAATGCTGTGGGTATCAACAGTGTGATGTGCTC 1420
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 73 CTACATCTGAGCCTTAATCAATGCTGTGGGTATCAACAGTGTGATGTGCTC 132
 QY 1421 ACAGTGTCTCAAGCACAAGAGCTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGG 1480
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 133 ACAGTGTCTCAAGCACAAGAGCTTTGTGTCCATGCTGTGTCTAGAAAACAGACTGGGG 192
 QY 1481 AACCTTATGTGAGCAGCAGATCCACAGTGAAGAGGTAAATGCTTCTTCTTCTT 1540
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 193 AACCTTATGTGAGCAGCAGATCCACAGTGAAGAGGTAAATGCTTCTTCTTCTT 252
 QY 1541 GATCTTCGTGTGTGGGAGACTTTCAGAGACTTTGTGGCTGTGAGGCTTATTAAGCAGCAG 1600
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 253 GATCTTCGTGTGTGGGAGACTTTCAGAGACTTTGTGGCTGTGAGGCTTATTAAGCAGCAG 312
 QY 1601 ACAGTGTCTCAAGTGAATGATTAATCAATGCTTCCCTGTCACATCTTCCCAAGGGGAATG 1660
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 313 ACAGTGTCTCAAGTGAATGATTAATCAATGCTTCCCTGTCACATCTTCCCAAGGGGAATG 372
 QY 1661 GATCTTTCACCAAGAGCTCACAGAGATTTTCCACAGAGATGC 1703
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 373 GATCTTTCACCAAGAGCTCACAGAGATTTTCCACAGAGATGC 415

Search completed: August 14, 2003, 20:11:36
 Job time : 7288 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:33:50 ; Search time 493.12 Seconds
(without alignments)
9271.801 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 2043
Sequence: 1 gaattcatgtgttggtta.....tgggacccaataaaaaaa 2043

Scoring table:
IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1504479 seqs, 1118970152 residues
Total number of hits satisfying chosen parameters: 3008958

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_NA:*

- 1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
- 2: /cgn2_6/ptodata/1/pubpna/PCr_NEW_PUB.seq:*
- 3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
- 4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
- 5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
- 6: /cgn2_6/ptodata/1/pubpna/PCrUS_PUBCOMB.seq:*
- 7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
- 8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
- 9: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 10: /cgn2_6/ptodata/1/pubpna/US09_PUBCOMB.seq:*
- 11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
- 12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
- 13: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 14: /cgn2_6/ptodata/1/pubpna/US10_PUBCOMB.seq:*
- 15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
- 16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
- 17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1784.8	87.4	2032	9 US-09-816-825-1	Sequence 1, Appl1
2	1784.8	87.4	2032	13 US-10-007-262-2	Sequence 194, Appl1
3	480.2	23.5	517	10 US-09-819-580-194	Sequence 194, Appl1
4	478.8	23.4	505	10 US-09-998-598-2595	Sequence 2595, Ap
5	412.8	20.2	2544	9 US-09-927-602-1	Sequence 1, Appl1
6	336.2	16.5	48436	9 US-09-927-602-1	Sequence 38, Appl1
7	294.4	14.4	389	10 US-09-919-580-264	Sequence 264, App
8	119.4	5.8	2393	9 US-09-833-790-244	Sequence 244, App
9	119.4	5.8	2409	14 US-10-212-933-3	Sequence 3, Appl1
10	108	5.3	2150	14 US-10-212-933-1	Sequence 1, Appl1
11	62.6	3.1	731	9 US-09-735-705-63	Sequence 63, Appl1
12	62.6	3.1	731	10 US-09-850-716A-63	Sequence 63, Appl1
13	62.6	3.1	731	10 US-09-897-718-63	Sequence 63, Appl1
14	62.6	3.1	731	11 US-09-466-396A-63	Sequence 63, Appl1
15	62.6	3.1	731	12 US-10-117-982-63	Sequence 63, Appl1
16	59	2.9	303	9 US-09-833-790-243	Sequence 243, App

17	59	2.9	303	9 US-09-833-790-401	Sequence 401, App
c 18	42.8	2.1	395	9 US-09-833-790-207	Sequence 207, App
c 19	42.4	2.1	2976	14 US-10-128-714-355	Sequence 355, App
c 20	42.4	2.1	2977	14 US-10-128-714-5355	Sequence 5355, App
21	41.6	2.0	720	13 US-10-027-632-22081	Sequence 22081, A
22	41.6	2.0	720	13 US-10-027-632-149086	Sequence 149086, A
23	41.4	2.0	1515	14 US-10-156-761-7394	Sequence 7394, Ap
24	41.4	2.0	9025608	14 US-10-156-761-1	Sequence 1, Appl1
25	40.2	2.0	1011	14 US-10-156-761-7058	Sequence 7058, App
c 26	40.2	2.0	9025608	14 US-10-156-761-1	Sequence 1, Appl1
27	40	2.0	2838	9 US-09-815-242-7888	Sequence 7888, App
28	39.6	1.9	344	9 US-09-864-761-19694	Sequence 19694, A
29	39.6	1.9	399	9 US-09-864-761-2913	Sequence 2913, Ap
30	39.6	1.9	738	14 US-10-156-761-1406	Sequence 1406, App
31	39.6	1.9	7092	14 US-10-037-270-252	Sequence 252, App
c 32	38.6	1.9	7269	14 US-10-037-270-251	Sequence 251, App
c 33	39.6	1.9	20165	13 US-10-159-330-7	Sequence 7, Appl1
c 34	39.6	1.9	20165	14 US-10-159-334-7	Sequence 7, Appl1
c 35	39.2	1.9	771	14 US-10-156-761-3385	Sequence 3385, Ap
36	38.2	1.9	1512	14 US-10-156-761-7404	Sequence 7404, Ap
c 37	38	1.9	666	13 US-10-027-632-222628	Sequence 222628, A
c 38	38	1.9	6381	12 US-10-311-455-939	Sequence 939, App
39	38	1.9	6381	12 US-10-240-485-71	Sequence 71, Appl1
40	37.8	1.9	963	14 US-10-156-761-5861	Sequence 5861, Ap
41	37.6	1.8	362	13 US-10-027-632-272279	Sequence 272279, A
c 42	37.6	1.8	1022	13 US-10-027-632-9247	Sequence 9247, Ap
c 43	37.6	1.8	1022	13 US-10-027-632-9248	Sequence 9248, Ap
c 44	37.6	1.8	6130	12 US-10-311-455-102	Sequence 102, App
c 45	37.6	1.8	7607	12 US-10-240-485-168	Sequence 168, App

ALIGNMENTS

RESULT 1
US-09-816-825-1
Sequence 1, Application US/09816825
Patent No. US20010051370A1
GENERAL INFORMATION:
APPLICANT: Bistrup, Annette
APPLICANT: Rosen, Steven D.
APPLICANT: Hemmerlich, Stefan
TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
FILE REFERENCE: 6510-107CON
CURRENT APPLICATION NUMBER: US/09/816, 825
PRIOR FILING DATE: 2001-03-22
PRIOR APPLICATION NUMBER: 09/045, 284
NUMBER OF SEQ ID NOS: 9
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2032
TYPE: DNA
ORGANISM: Homo sapiens
US-09-816-825-1

Query Match 87.4% Score 1784.8; DB 9; Length 2032;
Best Local Similarity 96.8%; Pred. No. 0;
Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

Qy	94	AAGCCCAAGCAGCAAGTCTTCACATTCAGCAATGCTACTGCTTAATAAATGAACT	153
Db	148	AAGCCGCTTGAAGCAAGCTTCACATTCAGCAATGCTACTGCTTAATAAATGAACT	207
Qy	154	CTGCTGTTTGTGTTTCCAGATGCGCATCTTGCTCTATTTCTTCACATGTACAGCA	213
Db	208	CTGCTGTTTGTGTTTCCAGATGCGCATCTTGCTCTATTTCTTCACATGTACAGCA	267
Qy	214	CAACATGCTCCCTGCTATGAAGGACACCCGAGGACGATGCTGTTCTG	273
Db	268	CAACATGCTCCCTGCTATGAAGGACACCCGAGGACGATGCTGTTCTG	327
Qy	274	TTCTGCGGCTCTGCGCTCTTTTGTGCGGACGCTTTTGGAGACACCGATGTTT	333

```

Db 328 TTCTGGGCTCTGGCTCTCTTGTGTGGGCGAGCTTTTGGGCGACCCAGCCATGTTT 397
QY 334 CTACCTGATGAGAGCCCGCTGGGACGCTGTGATGACCTTCAAGCAGAGCAGCCGCTGAT 393
Db 388 CTACCGATGAGAGCCCGCTGGGACGCTGTGATGACCTTCAAGCAGAGCAGCCGCTGAT 447
QY 394 GCTGCGACATGCTGTGCGGAGATCTGATAGGCGCGCTCTTCTTGTGCGACATGAGCTCT 453
Db 448 GCTGCGACATGCTGTGCGGAGATCTGATAGGCGCGCTCTTCTTGTGCGACATGAGCTCT 507
QY 454 TGAATGCTACATGGAAGCTGTGTCCCGGAGACAGTCCAGGCTCTTCAAGGAGAGAG 513
Db 508 TGAATGCTACATGGAAGCTGTGTCCCGGAGACAGTCCAGGCTCTTCAAGGAGAGAG 567
QY 514 CCGGGGCTGTGTGTGACAGTGTGACATGATCCAGCAAGATGAATATATCCCGG 573
Db 568 CCGGGGCTGTGTGTGACAGTGTGACATGATCCAGCAAGATGAATATATCCCGG 627
QY 574 GGGTCACTGAGGCTCTGTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 633
Db 628 GGGTCACTGAGGCTCTGTGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 687
QY 634 CTCTACAGCCAGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 693
Db 688 CTCTACAGCCAGTGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 747
QY 694 GCTGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 753
Db 748 GCTGCTGAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 807
QY 754 CCGTGTCCCTGTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 813
Db 808 CCGTGTCCCTGTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 867
QY 814 GGGGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 873
Db 868 GGGGAGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 927
QY 874 CTGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 933
Db 928 CTGCGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 987
QY 934 AGGCTACCTGCTGTGCGGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 993
Db 988 AGGCTACCTGCTGTGCGGCTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1047
QY 994 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1053
Db 1048 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1107
QY 1054 CACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1113
Db 1108 CACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1167
QY 1114 TGTCTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1173
Db 1168 TGTCTCCCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1227
QY 1174 CTGTGGGAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1233
Db 1228 CTGTGGGAGATGAGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1287
QY 1234 AAACCTGTGCTGATCTTCTGTCTACCTGAGCTGTCCGTGAGCAAAATCCAGTAAGAGG 1293
Db 1288 AAACCTGTGCTGATCTTCTGTCTACCTGAGCTGTCCGTGAGCAAAATCCAGTAAGAGG 1347
QY 1294 TTGAGAAAGCTTGTGCTGACAGCTGTGTGAGCTGATGATGATGATGATGATGATGAT 1353
Db 1348 TTGAGAAAGCTTGTGCTGACAGCTGTGTGAGCTGATGATGATGATGATGATGATGAT 1407
QY 1354 GCGTGTGCTGATCTGTGAGCTTACTGATGATGATGATGATGATGATGATGATGATGAT 1413

```

```

Db 1408 GCGTGTGCTGATCTGTGAGCTTAACTACAGTGTGTGGGATTCACAGCTGATGAGT 1467
QY 1414 TGTGTCCACAGCTGTCTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1473
Db 1468 TGTGTCCACAGCTGTCTACAGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1527
QY 1474 ACTGGGAACTTATGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1533
Db 1528 ACTGGGAACTTATGTGAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAGCAG 1587
QY 1534 TTTTCTTGATCTTCTGTGTGGGAGAGCTTACAGACTTGTGGGCTGGAGGCTTATTA 1593
Db 1588 TTTTCTTGATCTTCTGTGTGGGAGAGCTTACAGACTTGTGGGCTGGAGGCTTATTA 1647
QY 1594 GCACGACAGATATCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 1653
Db 1648 GCACGACAGATATCAGTGAATGATGATGATGATGATGATGATGATGATGATGATGATG 1707
QY 1654 GGGAAATGATCTTTCACCAAGAGCTCACCAGATTTTCCACAGAGATGGAATTTCTGAG 1713
Db 1708 GGGAAATGATCTTTCACCAAGAGCTCACCAGATTTTCCACAGAGATGGAATTTCTGAG 1767
QY 1714 CCCTTGAGTTCCTCAATGGGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1773
Db 1768 CCCTTGAGTTCCTCAATGGGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1826
QY 1774 ATGAGCTTGACAT-ACAGCTATCGGTAATCAGAAATATGAAACAAATCTCTG-ACAA 1831
Db 1827 ATGAGCTTGACATCAGAGCTATCGGTAATCAGAAATATGAAACAAATCTCTGACAA 1886
QY 1832 AGAGCAAGCTTAAATTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1890
Db 1887 AGAGCAAGCTTAAATTCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1946
QY 1891 CATTTTCCATTCATAGAAACTTGTACCTGTGAAGCTTGCATCTGTTAATACTTAA 1950
Db 1947 CATTTTCCATTCATAGAAAGACTTGTACCTGTGAAGC-TGCCATCTGTTAATACTTAA 2005
QY 1951 TTCCCAATTAAG 1962
Db 2006 TTCCCAATTAAG 2017

RESULT 2
US-10-007-262-2
; Sequence 2, Application US/10007262
; Publication No. US2002016478A1
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerlich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/10/007,262
; PRIOR FILING DATE: 2001-11-08
; PRIOR APPLICATION NUMBER: EARLIER APPLICATION NUMBER: 09/190,911
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens
US-10-007-262-2

Query Match 87.4%; Score 1784.8; DB 13; Length 2032;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

QY 94 AAGCCCAAGCACAAGAGCTTCTCAGCTTCAAGCAATGCTAGTCCCTTAAATGAAGCT 153
Db 148 AAGCCCGCTTTCAGAGGCTTTCACACTTTCAGCAGATGCTACTGCTTAAATGAAGCT 207

```

154 CCTGCTGTTCTGTTCCAGATGGCATCTTGCTATTTCTCCAGATGTACAGCA 213
208 CTTGCTGTTCTGTTCCAGATGGCATCTTGCTATTTCTCCAGATGTACAGCA 267
214 CAACATGAGTCCCTGCTATGAGGACACAGCCGAGCGCATGACGCTGTTCTGTC 273
268 CAACATGAGTCCCTGCTATGAGGACACAGCCGAGCGCATGACGCTGTTCTGTC 327
274 TTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 333
328 TTCTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 387
334 CTACCTGATGAGCCCGCTGAGCTGCTGATGACCTTCAAGAGAGACACCGCTGAT 393
388 CTACCTGATGAGCCCGCTGAGCTGCTGATGACCTTCAAGAGAGACACCGCTGAT 447
394 GCTGACATGAGCTGCTGCGGATGCTGATGAGGCGCTGCTGCTGCTGCTGCTGCT 453
448 GCTGACATGAGCTGCTGCGGATGCTGATGAGGCGCTGCTGCTGCTGCTGCTGCT 507
454 TGATGCGCATGAGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 513
508 TGATGCGCATGAGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 567
514 CCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 573
568 CCGGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 627
574 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 633
628 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 687
634 CTCTGACAGCATGAGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 693
688 CTCTGACAGCATGAGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 747
694 GCTGCTGAGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 753
748 GCTGCTGAGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 807
754 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 813
808 GCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 867
814 GGGGAGCATGAGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 873
868 GGGGAGCATGAGTGAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 927
874 CTGCGAAGCCAGCTGAGATCTACAGAACCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 933
928 CTGCGAAGCCAGCTGAGATCTACAGAACCATCATGCTGCTGCTGCTGCTGCTGCTGCTGCT 987
934 AGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 993
988 AGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1047
994 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1053
1048 AATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1107
1054 CACCGAGGCAAGGAGCATGAGTGAACAGCTTTCACACAAATGCGAGGATGCGCTTAA 1113
1108 CACCGAGGCAAGGAGCATGAGTGAACAGCTTTCACACAAATGCGAGGATGCGCTTAA 1167
1114 TGTCTCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1173
1168 TGTCTCCAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1227
1174 CTGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1233
1228 CTGAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1287

1234 AACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1293
1288 AACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1347
1294 TTGAGAAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1353
1348 TTGAGAAAGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1407
1354 GCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1413
1408 GCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1467
1414 TGTGTCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1473
1468 TGTGTCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1527
1474 ACTGGGAACTTATGATGAGAGACATCCACAGTGAAGAGGATGCTGCTTCTTC 1533
1528 ACTGGGAACTTATGATGAGAGACATCCACAGTGAAGAGGATGCTGCTTCTTC 1587
1534 TTTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1593
1588 TTTTCTGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1647
1594 GCACGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1653
1648 GCACGACACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1707
1654 GGGAT 1713
1708 GGGAT 1767
1714 CCGTTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1773
1768 CCGTTGGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1826
1774 ATGAGCTTGACCAT-ACAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1831
1827 ATGAGCTTGACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1886
1832 AGAGCAAGCTTAAAGTTACACAGAGTGGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 1890
1887 AGAGCAAGCTTAAAGTTACACAGAGTGGCTGCTGATGATGATGATGATGATGATGATGATGATGAT 1946
1891 CATTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1950
1947 CATTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2005
1951 TTCCCAATTAAG 1962
2006 TTCCCAATTAAG 2017

RESULT 3
US-09-919-580-194
; Sequence 194, Application US/09919580
; Patent No. US20020110832A1
; GENERAL INFORMATION:
; APPLICANT: Pyle, Ruth
; APPLICANT: Xu, Jianshun
; APPLICANT: Secret, Heather
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
; FILE REFERENCE: 210121.552
; CURRENT APPLICATION NUMBER: US/09/919,580
; NUMBER OF SEQ ID NOS: 934
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 194
; LENGTH: 517
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-919-580-194

Query Match 23.5%; Score 480.2; DB 10; Length 517;
Best Local Similarity 99.0%; Pred. No. 5.4e-141;
Matches 504; Conservative 0; Mismatches 3; Indels 2; Gaps 2;

QY 1298 GAAGGCTTGGTCCACCGCTGGTGTACGCTCAGTACTTCTGTGATGCTTGTGAGCCT 1357
DB 10 GAGGGCTTGGTCCACCGCTGGTGTACGCTCAGTACTTCTGTGATGCTTGTGAGCCT 69
QY 1358 TCCCTACATCTGTGAGCCTTAACATGCTGTGGGTATACACAGTGTGAGTGTG 1417
DB 70 TCCCTACATCTGTGAGCCTTAACATGCTGTGGGTATACACAGTGTGAGTGTG 129
QY 1418 TCCACAGTGTCAAGCAGCAGAGACTTGTGTCCATGCTGTGTCTAGAAAAACAGATG 1477
DB 130 TCCACAGTGTCAAGCAGCAGAGACTTGTGTCCATGCTGTGTCTAGAAAAACAGATG 189
QY 1478 GGGAACTTATGTGAGCAGCAGATCCACAGTGAAGAGGTATGCTCTCTCTTTT 1537
DB 190 GGGAACTTATGTGAGCAGCAGATCCACAGTGAAGAGGTATGCTCTCTCTTTT 249
QY 1538 CTGTATCTTCTGTGCGGAGACTTACAGACTTGTGGCTGGAGGCTATTAGCAC 1597
DB 250 CTGTATCTTCTGTGCGGAGACTTACAGACTTGTGGCTGGAGGCTATTAGCAC 309
QY 1598 GACACAGTATCAAGTGAATGATCCATAAACCCTCCCTGCACATCTTCCCAATGGGA 1657
DB 310 GACACAGTATCAAGTGAATGATCCATAAACCCTCCCTGCACATCTTCCCAATGGGA 369
QY 1658 ATGATCTTTCACAAAGAGCTCAGCAGATTTTCACAGAGATGGAATTCGAGCCT 1717
DB 370 ATGATCTTTCACAAAGAGCTCAGCAGATTTTCACAGAGATGGAATTCGAGCCT 429
QY 1718 TGAAGTTCCTCATGGATTCAGAGAAAGTGGGAACAGGTGGATGCTACTATGA 1777
DB 430 TGAAGTTCCTCATGGATTCAGAGAAAGTGGGAACAGGTGGATGCTACTATGA 488
QY 1778 GCTTGACAT-ACAGTATCGTAAACG 1805
DB 489 GCTTGACATCAACAGTATCGTAAACG 517

RESULT 4
US-09-998-598-2595
; Sequence 2595, Application US/09998598
; Patent No. US20020150922A1
GENERAL INFORMATION:
APPLICANT: Stolk, John A.
APPLICANT: Xu, Jianshun
APPLICANT: Chenault, Ruth A.
APPLICANT: Meagher, Madelein Joy
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY AND
TITLE OF INVENTION: DIAGNOSIS OF COLON CANCER
FILE REFERENCE: 210121.561
CURRENT APPLICATION NUMBER: US/09/998,598
CURRENT FILING DATE: 2001-11-16
NUMBER OF SEQ ID NOS: 2606
SOFTWARE: Corixa Invention Disclosure Database
SEQ ID NO 2595
LENGTH: 505
TYPE: DNA
ORGANISM: Homo sapiens
US-09-998-598-2595

Query Match 23.4%; Score 478.8; DB 10; Length 505;
Best Local Similarity 99.2%; Pred. No. 1.5e-140;
Matches 502; Conservative 0; Mismatches 2; Indels 2; Gaps 2;

QY 1301 GGCCTTGTGCGACCGCTGTGAGCCTCAGTCACTTCTGTGAATGCTTGTGAGCCTTC 1360
DB 1 GGCCTTGTGCGACCGCTGTGAGCCTCAGTCACTTCTGTGAATGCTTGTGAGCCTTC 60
QY 1361 CTACATCTGTGAGCCTTAACATGCTGTGGGTATCACACTGAGTGTGCTCC 1420

DB 61 CTACATCTGTGAGCCTTAACATGCTGTGGGTATCACACTGAGTGTGCTCC 120
QY 1421 ACACGCTGCACGACAGAGAGACTTGTGTCCATGCTGTGTCTAGAAAAACAGACTGGG 1480
DB 121 ACACGCTGCACGACAGAGAGACTTGTGTCCATGCTGTGTCTAGAAAAACAGACTGGG 180
QY 1481 AACCTTATGTGAGCAGCAGATCCACAGTGAAGAGGTATGCTCTCTCTCTTCTT 1540
DB 181 AACCTTATGTGAGCAGCAGATCCACAGTGAAGAGGTATGCTCTCTCTCTTCTT 240
QY 1541 GATCTTCTGTGTGGGACAGTTCAGAGACTTGTGTGCTGTGAGGCTATTAGCAGC 1600
DB 241 GATCTTCTGTGTGGGACAGTTCAGAGACTTGTGTGCTGTGAGGCTATTAGCAGC 300
QY 1601 ACAGTATCAAGTGAATGATTCATTAACCTCCCTGCACATCTTCCCAATGGGGAAG 1660
DB 301 ACAGTATCAAGTGAATGATTCATTAACCTCCCTGCACATCTTCCCAATGGGGAAG 360
QY 1661 GATCTTTCACCAAGAGCTCACAGCATTTCACAGAGATGCAATTCGAGCCTTGG 1720
DB 361 GATCTTTCACCAAGAGCTCACAGCATTTCACAGAGATGCAATTCGAGCCTTGG 420
QY 1721 AGTTCCTCAATGGGATTCAGAGAGAGTGGGAACAGGTGGATGCTACTTATGAGCT 1780
DB 421 AGTTCCTCAATGGGATTCAGAGAGAGTGGGAACAGGTGGATGCTACTTATGAGCT 479
QY 1781 TGACAT-ACAGTATCGTAAACG 1805
DB 480 TGACATCAACAGTATCGTAAACG 505

RESULT 5
US-09-927-602-1
; Sequence 1, Application US/09927602
; Patent No. US20020061562A1
GENERAL INFORMATION:
APPLICANT: Fukuda, Michiko N.
APPLICANT: Akama, Tomoya O.
TITLE OF INVENTION: Methods of Treating Macular Corneal
TITLE OF INVENTION: Dystrophy
FILE REFERENCE: P-LJ 4852
CURRENT APPLICATION NUMBER: US/09/927,602
CURRENT FILING DATE: 2001-08-09
PRIOR APPLICATION NUMBER: US 09/638,211
PRIOR FILING DATE: 2000-08-11
NUMBER OF SEQ ID NOS: 38
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 1
LENGTH: 2544
TYPE: DNA
ORGANISM: Homo Sapien
FEATURE:
NAME/KEY: CDS
LOCATION: (693)...(1877)
US-09-927-602-1

Query Match 20.2%; Score 412.8; DB 9; Length 2544;
Best Local Similarity 64.4%; Pred. No. 3.8e-119;
Matches 653; Conservative 0; Mismatches 352; Indels 9; Gaps 2;

QY 248 GAGCGCATGACGCTGTGCTGTCTGTGCGCTGTGCTCTCTTTTGTGGGCGAG 307
DB 807 GCGCGCATGACGCTGTGCTGTCTGTGCGCTGTGCGCTGTGCTCTCTTTTGTGGGCGAA 866
QY 308 CTTTGTGGGACAGCCAGATTTTCTACGATGAGAGAGCCGCGCTGAGAGTGTGAGAG 367
DB 867 CTTTGTGGGACAGCCAGATTTTCTACGATGAGAGAGCCGCGCTGAGAGTGTGAGAG 926
QY 368 ACCTTCAGCAGAGACCGCTGTGATGCGCATGAGCTGTGCGGATGTGATGAGGCGC 427
DB 927 ACCCTGTGCGAGGAGCGCCCGCAACGCTGTGACATGTGCTGTGCGGAGCTGTGCTCC 986

OY	428	GTCTCTTGTCGCGACATGAGCGCTGTTTGATGCTCAACATGGAACCGTGGTCCCGGAGACAG	487
Db	987	GTCTTCTCTGTCGACATGAGCGTGTGATGCTTATCTG---CCTTGGCGCCGAACCTG	1043
OY	488	TCGAGCGCTTTTCAGTGGGAGAACAGCCGGGCGCTGTGCTGTCGACCTGACTGTGACATC	547
Db	1044	TCGACGCTTTCCAGTGGGCGGTGAGCGCGTGCACATGTGTGTGCGCACCGCGCTGCAATGTC	1103
OY	548	ATCCACACAGATGAAATCATCCCGGGGCTCAGTCGAGGCTCTGTGCAAGTCAACAGCC	607
Db	1104	TTTTCGCCGAGGCGGCATCAGCAGGAGGCGCTGTGTGCAAGCCACACTGTGGCGCCGACATCC	1163
OY	608	TTTAGGAGTGGGAGAAAGGCGCTGGCGCTCCACAGACACAGTGGTGTCAAGAGGTGGC	667
Db	1164	TTTCAACCTGGCGCGGAGGCGCTGGCGCTCTTACAGCCACAGTGGTGTCAAGAGGTGGC	1223
OY	668	TTTCTTCAACCTGCAAGTCCCTCTTACCCGCTGTGTAAGACCCCTCTCTACACTGCATATC	727
Db	1224	TTTCTTCAACCTGCAAGTGGCTGTACCCGCTGTGTCAAGCACCCCGGGCTCTAACCTTACGATC	1283
OY	728	GTGCACTGGTTCGCGGAGCCCGCGCGGTTCCTGTTCCGAGAACGCAAGAGGAGAT	787
Db	1284	GTGCACTGGTGGCGGACCCGCGCGGTGTGCTGCCGCGGAGACAGACAGCCAAAGCT	1343
OY	788	CTCATGATTTGACATCGCATGTGTATGGGGGAGCAGATGAGCAAAAACCTCAAGAGAGAC	847
Db	1344	CTGGCGGCTGACCAACGCGCATGTGTGTGGACACCAACGCGACGTGGTGGAGGCCGACCC	1403
OY	848	CAACCTCATATGTGATGTGAGTATCTGCGCAAGCCAGCTGAGATCTTACAAACCATC	907
Db	1404	GGCCTGGCGGTGGTGGCGGAGGTGTGCCGTGAGCCAGTACGATCGCCGAGGCGGCACA	1463
OY	908	CAGTCTTGGCCCAAGGCGCTGCAGGAACGCTTACTGCTTGTGGCGTATGAGAGCTGGCT	967
Db	1464	CTCAAGCGCGCACCCCTTCTGTGCGGCGCGCTACCGCGCTGGTGGCGCTTTCGAGAGCACTGGCG	1523
OY	968	CGACCGCCCTGGCGCCACAGACTTCCCGAATGTATGAAATGCGGGATTTGTAATCTTGGCC	1027
Db	1524	CGGAGCGCGCTGCGAGAAATCCGTGCGCTCTTACGCTTTCACCTAGGCTCAGTCTCACGCTCA	1583
OY	1028	CATCTTCAGACCTGGGTGCATATACATCACCCGAGGCAAGGGCATGGGTG-----ACCAC	1081
Db	1584	CAGCTCGAGGCGCTGGATCCATATACATCACCCAGCGATGTGAGACCTGGTGGCGCGCGCAA	1643
OY	1082	GCTTTCACACAAATGCCAGGAGTCCCTTAATGTCTCCACAGGCTTGGCGGTGCTTTTG	1141
Db	1644	GCTTTCAAAGATCTGTCTCAGGAATGCGCTCAACGCTCTCCAGGCGCTGGCGCATATGCGCTG	1703
OY	1142	CCCATGAAAGGTTTCTCGACTTCAGAAACCTGTGGCATTTGCCATGAAATTTGCTGGGC	1201
Db	1704	CCCTTTGCGCAAGATCCGCGCGGTGCAGGAACTGTGGCTGTGGGTGCACCTCTGGGCG	1763
OY	1202	TACGCGCACGTGCAGATCTGAACACAGAAACAGAAACCTGTTGCTGATCTTCTG	1255
Db	1764	TACGCGCGCTGTGTACTGTGAGAGAGAGACGCGCAACTCTGGCCCTTGTATCTGTG	1817
RESULT 6			
US-09-927-602-38			
Sequence 38, Application US/09927602			
Patent No. US2002061562A1			
GENERAL INFORMATION:			
APPLICANT: Fukuda, Michiko N.			
APPLICANT: Akama, Tomoya O.			
TITLE OF INVENTION: Methods of Treating Macular Corneal			
TITLE OF INVENTION: Dystrophy			
FILE REFERENCE: P-LJ 4852			
CURRENT APPLICATION NUMBER: US/09/927.602			
CURRENT FILING DATE: 2001-08-09			
PRIOR APPLICATION NUMBER: US 09/638,211			
PRIOR FILING DATE: 2000-08-11			
NUMBER OF SEQ ID NOS: 38			
SOFTWARE: FastSeq for Windows Version 4.0			

Query Match	16.5%	Score 336.2	DB 9	Length 48436
Best Local Similarity	65.0%	Pred. No. 5.7e-94		
Matches 513	Conservative 0	Mismatches 273	Indels 3	Gaps 1
OY	248	GAGCGCAATGACACGCTGCTGTTCTCTCTCCGTGCGCTCTGGCTCTTCTTTGTTGGGGCAG	307	
Db	47651	GGGGGCGATGATGCTGCTGCTCTCTCTCTGGGCGCTGGGCTGCTCTCTGGGGCAG	47710	
OY	308	CTTTTGGGCGACACCCAGATGTTTCTTACCTGATGAGCCCGCTGGACGCTGTGGATG	367	
Db	47711	CTTTTCAACCGACCGCCGACGCTCTTCTTACCTGATGAGCCCGCTGGACGCTGTGGATG	47770	
OY	368	ACCTTCAAGCAGACACCGGCTGGATGCTGACATGCTGTGGGGATGCTGATGAGGGCC	427	
Db	47771	ACCTGTGCGACGGGCGACCGCGCAACGCTGACATGCTGTGGGGATGCTGATGAGGGCC	47830	
OY	428	GCTCTTCTGTGCGACATGAGCGTCTTTGATGCTTACATGAACTGTGCTCCCGAGACAG	487	
Db	47831	GCTCTCTGTGCGACATGAGCGTCTTTGATGCTTACATGAACTGTGCTCCCGAGACAG	47880	
OY	488	TCCACGCTCTTTCAGTGGGAGAACGCGGGCCCTGTGTTCTGACACTGCTGTGACATC	547	
Db	47888	TCCACGCTCTTTCAGTGGGAGAACGCGGGCCCTGTGTTCTGACACTGCTGTGACATC	47940	
OY	548	ATCCACAAAGTAAATCATCCCGGGGCTCACTGCGAGGCTCCGTGACATCAACAGCC	607	
Db	47948	TTTCCCGGAGGCGCATATGAGAGGAGCGCGTGTGCAAGCACTGTGCGCGCGAGTCC	48007	
OY	608	TTTGAAGTGTGAGAGAGCGCTGCGGCTCTTACACGACCGTGTGCTCAAGAGAGTGGC	667	
Db	48008	TTTCAACGCTGGCGGGAGGCGCTGCGGCTCTTACACGACCGTGTGCTCAAGAGAGTGGC	48060	
OY	668	TTTCTTCAACCTGCACTGCTCTTACCGCGCTGCTGAAGACCCCTTCCCTCAACTGTGATATC	727	
Db	48068	TTTCTTCAACCTGCACTGCTCTTACCGCGCTGCTGAAGACCCCGCTCAACTGTGATATC	48122	
OY	728	GTCGACCTGGGCGGGGACCGCGGCGCTGCTCCGAGAAACGCAAAAGGAGAT	787	
Db	48128	GTCGACCTGGGCGGGGACCGCGGCGCTGCTCCGAGAGAGAGCAAGGAT	48180	
OY	788	CTCATGATGACATGCGCATTTGTATGGGCGAGCATGAGCAAAAATCTCAAGAGAGAGAC	847	
Db	48188	CTGGCGGCTGCAACAGCGCATGCTGTGGGCAACAGGCGACGTGGGAGGCGGACCC	48240	
OY	848	CAACCCATGATGATGATGACAGGTATCTGCGCAAGCGAGCTGAGATCTCAAGACCATC	907	
Db	48248	GGCGTGGCGGCTGTGGGAGGTGTGCGCTGATGACCATGCGCGGAGGCGCGCAC	48300	
OY	908	CAGTCTTGGCCCAAGCCCTGTGAGAAACGCTACTGCTTGTGGCTATGAGAGCTGGCT	967	
Db	48308	CTCAAGCGCGCACCCCTTGTGTGCGGCGGCGCTACCGCTGTGGCTGTGAGGACCTGGCG	48360	
OY	968	CGAGCGCCCTGTGGCGCCAGACTTCCCGAATGATGAATTCGTGGGATGGAATCTTGGCC	1027	
Db	48368	CGGAGCGCGCTGTGGGAGAAATCTGCGGCTTACCGCTCACTGGGCTCACTGTACAGCCA	48420	
OY	1028	CATCTTCAG	1036	
Db	48428	CAGCTCGAG	48436	

RESULT 7

US-09-919-580-264

Sequence 264, Application us/09919580

Patent No. US20020110832A1

GENERAL INFORMATION:

APPLICANT: Pyle, Ruth

LOCATION: 719, 723, 725, 730, 731
OTHER INFORMATION: n = A,T,C or G
US-10-117-982-63

Query Match 3.1%; Score 62.6; DB 12; Length 731;
Best Local Similarity 58.2%; Pred. No. 1.3e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY	648	TGGTCTCAAGGAGGTGCGCTTCTTCACCTGCAGTCCCTACCCGCTGTGAAAGACC	707
DB	3	TAGTCATAAAGGGGTGCGCGTCTTCGACGTGGCGGCTTGGCCGCACTGTCGAGACC	62
QY	708	CCTCCCTCAACCTGCATATCGTGCACCTGTCCGGACCCCGGGCCCGTTCCTCC	767
DB	63	CGCCCTGAGCTCAAGGTGATCACTTGTGCTGATCCCGCGGGTGGCGAGTTTAC	122
QY	768	GAGAAAGCACAAGGAGATCTCATGATGACAGTCGATGTGATGGGGCAGCATGAGC	827
DB	123	GGATCCGCTCGGGCCACGGGCTCATCCGTGAGAGGCTTACAGGTGGTGGCAGCCGAGACC	182
QY	828	AAAAACTCA	836
DB	183	GCGAGCTCA	191

Search completed: August 14, 2003, 18:00:55
Job time : 513.12 secs

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:31:05 ; Search time 131.773 Seconds
(without alignments)
6843.163 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 2043
Sequence: 1 gaattccattgttggta.....tgggacctaataaaaaaa 2043

Scoring table: IDENTITY_NNC
Gapop 10.0 , Gapext 1.0

Searched: 569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters: 1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Issued_Patents_NA: *
1: /cgn2_6/ptodata/1/lna/5A.COMB.seq: *
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq: *
3: /cgn2_6/ptodata/1/lna/6A.COMB.seq: *
4: /cgn2_6/ptodata/1/lna/6B.COMB.seq: *
5: /cgn2_6/ptodata/1/lna/PCUTS.COMB.seq: *
6: /cgn2_6/ptodata/1/lna/backfiles1.seq: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query	Match Length	ID	Description
1	1813	88.7	2065	US-09-786-240-26	Sequence 26, Appl
2	1784.8	87.4	2032	US-09-045-284A-1	Sequence 1, Appl
3	1784.8	87.4	2032	US-09-190-911-2	Sequence 2, Appl
4	141.6	6.9	2156	US-08-899-514-1	Sequence 1, Appl
5	135.2	6.6	2190	US-09-015-188-1	Sequence 1, Appl
6	128.4	6.3	2354	US-08-655-878-1	Sequence 1, Appl
7	119.4	5.8	2409	US-09-263-023-3	Sequence 3, Appl
8	119.4	5.8	2409	US-09-471-867-3	Sequence 3, Appl
9	108	5.3	2150	US-09-263-023-1	Sequence 1, Appl
10	108	5.3	2150	US-09-471-867-1	Sequence 1, Appl
11	62.6	3.1	731	US-09-040-984-63	Sequence 63, Appl
12	62.6	3.1	731	US-09-123-912-63	Sequence 63, Appl
13	62.6	3.1	731	US-09-643-597-63	Sequence 63, Appl
14	62.6	3.1	731	US-09-480-884A-63	Sequence 63, Appl
15	62.6	3.1	731	US-09-542-615A-63	Sequence 63, Appl
16	62.6	3.1	731	US-09-606-421B-63	Sequence 63, Appl
17	42.2	2.1	570	US-09-252-991A-14204	Sequence 14204, A
18	42.2	2.1	759	US-09-252-991A-14242	Sequence 14242, A
19	40	2.0	4673	US-07-638-431-1	Sequence 1, Appl
20	40	2.0	4673	PCT-US92-00018-1	Sequence 1, Appl
21	39.6	1.9	7092	US-09-620-312D-252	Sequence 252, App
22	39.6	1.9	7269	US-09-620-312D-251	Sequence 251, App
23	39.6	1.9	20165	US-09-609-816-7	Sequence 7, Appl
24	38.4	1.9	3060	US-09-252-991A-4686	Sequence 4686, Ap
25	38.4	1.9	3455	US-09-252-991A-4757	Sequence 4757, Ap
26	38.2	1.9	2367	US-09-252-991A-11973	Sequence 11973, A
27	38.2	1.9	3264	US-09-252-991A-12013	Sequence 12013, A

28	38.2	1.9	3279	US-09-252-991A-11905	Sequence 11905, A
29	37.8	1.9	498	US-09-252-991A-5845	Sequence 5845, Ap
30	37.8	1.9	1005	US-09-252-991A-5806	Sequence 5806, Ap
31	37.4	1.8	509	US-09-594-506-9	Sequence 9, Appl
32	37.2	1.8	612	US-09-252-991A-711	Sequence 711, App
33	37.2	1.8	1488	US-09-252-991A-623	Sequence 623, App
34	37	1.8	37	US-09-045-284A-7	Sequence 7, Appl
35	37	1.8	37	US-09-190-911-7	Sequence 7, Appl
36	37	1.8	1965	US-09-178-252-26	Sequence 26, Appl
37	36.8	1.8	330	US-09-252-991A-15704	Sequence 15704, A
38	36.8	1.8	768	US-09-252-991A-15652	Sequence 15652, A
39	36.8	1.8	825	US-09-252-991A-15741	Sequence 15741, A
40	36.8	1.8	1446	US-09-252-991A-7937	Sequence 7937, Ap
41	36.8	1.8	1767	US-09-252-991A-7614	Sequence 7614, Ap
42	36.8	1.8	2724	US-09-252-991A-7543	Sequence 7543, Ap
43	36.8	1.8	5894	US-08-665-259-24	Sequence 24, Appl
44	36.8	1.8	5894	US-08-762-500-24	Sequence 24, Appl
45	36.8	1.8	6525	US-08-762-500-74	Sequence 74, Appl

ALIGNMENTS

```
RESULT 1
US-09-786-240-26
Sequence 26, Application US/09786240
Patent No. 6558935
GENERAL INFORMATION:
APPLICANT: INCYTE PHARMACEUTICALS, INC.
APPLICANT: TANG, Y. Tom
APPLICANT: CORLEY, Neil C.
APPLICANT: GUEGLER, Karl J.
APPLICANT: BAUGHN, Mariah R.
APPLICANT: YUE, Henry
APPLICANT: HILLMAN, Jennifer L.
APPLICANT: AZIMZAI, Yalda
TITLE OF INVENTION: HUMAN TRANSFERASE PROTEINS
FILE REFERENCE: PR-0592 PCT
CURRENT APPLICATION NUMBER: US/09/786,240
CURRENT FILING DATE: 2002-03-12
PRIOR APPLICATION NUMBER: 09/150,657; unassigned; 09/186,779; unassigned; 60/133,6
PRIOR FILING DATE: 1998-09-10; 1998-09-10; 1998-11-04; 1998-11-04; 1999-05-11
NUMBER OF SEQ ID NOS: 33
SOFTWARE: PERL Program
SEQ ID NO 26
LENGTH: 2065
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: misc-feature
OTHER INFORMATION: incyte ID No. 6558935 2617407CB1
US-09-786-240-26
Query Match
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1902; Conservative 0; Mismatches 20; Indels 7; Gaps 7;
QY 94 AAGCCAGCCAGCAGAGCTCTTCACCTTCAGCAGCATGCTACTGCTTAAAGAAAGCT 153
Db 140 AAGCCCTCTTGGCAGAGCTCTTCACCTTCAGCAGCATGCTACTGCTTAAAGAAAGCT 199
QY 154 CCGCTCTTCTGCTTCCAGATGGCAGCATGCTTCTTCACATGTCAGCA 213
Db 200 CCGCTCTTCTGCTTCCAGATGGCAGCATGCTTCTTCACATGTCAGCA 259
QY 214 CAACATAGCTCCCTGCTCTATGAGGACAGCCGACGATGACGCTGCTTCTGCT 273
Db 260 CAACATAGCTCCCTGCTCTATGAGGACAGCCGACGATGACGCTGCTTCTGCT 319
QY 274 TTCTGCGCTCTGCTCTCTCTTTTGGGGGACGCTTTTGGGACGACCCAGATGTTT 333
Db 320 TTCTGCGCTCTGCTCTCTCTTTTGGGGGACGCTTTTGGGACGACCCAGATGTTT 379
```

QY 334 CTACCTGATGAGAGCCCGCTGGACGCTGTGATGACCTTCAAGCAGACGACCGCTGAT 393
 DB 380 CTACCTGATGAGAGCCCGCTGGACGCTGTGATGACCTTCAAGCAGACGACCGCTGAT 439
 QY 394 GGTGACATGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 453
 DB 440 GGTGACATGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 499
 QY 454 TATGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 513
 DB 500 TATGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 559
 QY 514 CCGGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 573
 DB 560 CCGGCGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 618
 QY 574 GGTGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 632
 DB 619 GGTGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 678
 QY 633 GGTGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 692
 DB 679 GGTGCTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 738
 QY 693 CCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 752
 DB 739 CCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 798
 QY 753 CCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 812
 DB 799 CCGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 858
 QY 813 TGGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 872
 DB 859 TGGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 918
 QY 873 TGGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 932
 DB 919 TGGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 978
 QY 933 AACGCTACCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 992
 DB 979 AACGCTACCTGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1038
 QY 993 GAATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1052
 DB 1039 GAATGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1098
 QY 1053 TCACCCGAGGCAAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1112
 DB 1099 TCACCCGAGGCAAGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1158
 QY 1113 ATGTCTCCAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1172
 DB 1159 ATGTCTCCAGGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1218
 QY 1173 CCTGTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1232
 DB 1219 CCTGTGGCGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1278
 QY 1233 GAAACCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1292
 DB 1279 GAAACCTGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1338
 QY 1293 GTTGAAGAGCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1352
 DB 1339 GTTGAAGAGCTTTTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1398
 QY 1353 AGCCTTGTCTATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1412
 DB 1399 AGCCTTGTCTATCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1458

QY 1413 TTGTCTCCACAGCTGCTCAAGCAGAAAGACTTTTGTGTCTCATGCTTGTCTTGAAGAA 1472
 DB 1459 TTGTCTCCACAGCTGCTCAAGCAGAAAGACTTTTGTGTCTCATGCTTGTCTTGAAGAA 1518
 QY 1473 GACTGGGAAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1532
 DB 1519 GACTGGGAAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1578
 QY 1533 CTTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1592
 DB 1579 CTTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1638
 QY 1593 AGCAGACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1652
 DB 1639 AGCAGACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1698
 QY 1653 GGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1712
 DB 1699 GGGGAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1758
 QY 1713 GGCCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1772
 DB 1759 GGCCTTGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1817
 QY 1773 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1830
 DB 1818 TATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1877
 QY 1831 AAGAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1889
 DB 1878 AAGAGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1937
 QY 1890 GCATTTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1949
 DB 1938 GCATTTTCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1996
 QY 1950 ATTCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2009
 DB 1997 ATTCCCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 2056
 QY 2010 AATGTTTAT 2018
 DB 2057 AATGTTTAT 2065

RESULT 2
 US-09-045-284A-1
 : Sequence 1, Application US/09045284A
 : Patent No. 6265192
 : GENERAL INFORMATION:
 : APPLICANT: Bistrup, Annette
 : APPLICANT: Rosen, Steven D.
 : APPLICANT: Hemmerlich, Stefan
 : TITLE OF INVENTION: GLYCOXYL SULFOTRANSFERASE-3
 : FILE REFERENCE: 6510-107051
 : CURRENT APPLICATION NUMBER: US/09/045, 284A
 : NUMBER OF SEQ ID NOS: 9
 : SOFTWARE: FastSeq for Windows Version 4.0
 : SEQ ID NO 1
 : LENGTH: 2032
 : TYPE: DNA
 : ORGANISM: Homo sapiens
 : US-09-045-284A-1

Query Match 87.4%; Score 1784.8; DB 3; Length 2032;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

QY 94 AAGCCCAAGCAGAGCTTCTTCCATGACGACAAATGCTTACCTTAAATAAATAAGCT 153
 DB 148 AAGCCCGCTTGCAGAGCTTCTTCCATGACGACAAATGCTTACCTTAAATAAATAAGCT 207

OY	154	CTCTGCTTTTGGTTTTCCAGATGGGCATCTTGGCTATCTTCCACATGTACAGCA	213
Db	208	CTCTCTGTTTTTGGTTTTCCAGATGGGCATCTTGGCTATCTTCCACATGTACAGCA	267
OY	214	CAACATCAGCTCCCTGTCTATGAAGGCAACACCAGAGGCATGTGACGGTGGTTCTCTC	273
Db	268	CAACATCAGCTCCCTGTCTATGAAGGCAACACCAGAGGCATGTGACGGTGGTTCTCTC	327
OY	274	TTTCTGGCGCTCTGGCTCTTCTTTTGTGGGGAGCCYTTTTGGGACGCCACCCAGATTTTT	333
Db	328	TTTCTGGCGCTCTGGCTCTTCTTTTGTGGGGAGCCYTTTTGGGACGCCACCCAGATTTTT	387
OY	334	CTACCTCATGGAGCCCGCTGGACACGTGTGGATACCTTCAAGACAGACACCCTGTGAT	393
Db	388	CTACCTCATGGAGCCCGCTGGACACGTGTGGATACCTTCAAGACAGACACCCTGTGAT	447
OY	394	GCTGCACATGCGCTGTGGGGATCTGATACGGGGCGTCTTCTGTGGCAGATGAGCGCTT	453
Db	448	GCTGCACATGCGCTGTGGGGATCTGATACGGGGCGTCTTCTGTGGCAGATGAGCGCTT	507
OY	454	TGATGCTCACTGNAACCTGCTCCCCGGAGACAGTCCAGCCTTTTTCAGTGGAGAACAG	513
Db	508	TGATGCTCACTGNAACCTGCTCCCCGGAGAGACAGTCCAGCCTTTTTCAGTGGAGAACAG	567
OY	514	CCGGGCCCCGTGTTCTGCACTGCGCTGTGATCATCCACAATATGAATCAATCCCCG	573
Db	568	CCGGGCCCCGTGTTCTGCACTGCGCTGTGATCATCCACAATATGAATCAATCCCCG	627
OY	574	GGCTCACTGCAGGGTCTCTGTGCAGTCAACACGCCCTTGTAGGTGGTGGAGAGGCTCGC	633
Db	628	GGCTCACTGCAGGGTCTCTGTGCAGTCAACACGCCCTTGTAGGTGGTGGAGAGGCTCGC	687
OY	634	CTCTTAACGCCAGTGTGTCTCAAGAGAGTGCCTTCTTCAACCTGCAGTCCCTCTACCC	693
Db	688	CTCTTAACGCCAGTGTGTCTCAAGAGAGTGCCTTCTTCAACCTGCAGTCCCTCTACCC	747
OY	694	GCTGCTGAAACACCCCTCCCTCAACCTGTGCATATGTGTGACCTGGTCCGGGACCCCCGGCG	753
Db	748	GCTGCTGAAACACCCCTCCCTCAACCTGTGCATATGTGTGACCTGGTCCGGGACCCCCGGCG	807
OY	754	CGTGTCCGTTCCCGAGAACGCACAAAGAGAGATCTCATGATTGCATGTCCATTTGTAT	813
Db	808	CGTGTCCGTTCCCGAGAGAACGCACAAAGAGAGATCTCATGATTGCATGTCCATTTGTAT	867
OY	814	GGGGCAGCATGACAAAAATCTCAAGAGAGAGACCAACCTACTATGTGTGATGCAGGTAT	873
Db	868	GGGGCAGCATGACAAAAATCTCAAGAGAGAGACCAACCTACTATGTGTGATGCAGGTAT	927
OY	874	CTGGCAAAAGCCAGCTGGAGATCTCAAGAACCATCCAGTCTTGGCCAAAGGCCCTGCAGGA	933
Db	928	CTGGCAAAAGCCAGCTGGAGATCTCAAGAACCATCCAGTCTTGGCCAAAGGCCCTGCAGGA	987
OY	934	ACGCTACCTGTTGTGCGCTATGTAGAGACCTGGCTGAGCCCCCTGTGGGCCACGATTTCCG	993
Db	988	ACGCTACCTGTTGTGCGCTATGTAGAGACCTGGCTGAGCCCCCTGTGGGCCACGATTTCCG	1047
OY	994	AATGTATGAATTCGTGGGATTGGAATTCCTGCCCCCATCTTCAAGACCTGGGTGATTAACAT	1053
Db	1048	AATGTATGAATTCGTGGGATTGGAATTCCTGCCCCCATCTTCAAGACCTGGGTGATTAACAT	1107
OY	1054	CACCCGAGGCAAGGGCATGGGTGACACACGCTTTCACACAAATGCGAGGATGCCCTTAA	1113
Db	1108	CACCCGAGGCAAGGGCATGGGTGACACACGCTTTCACACAAATGCGAGGATGCCCTTAA	1167
OY	1114	TGTCTCCAGGCTTGGCGCTGTGCTTTTGGCCCTATGAAGAAAGTTCTGCATCTCGAAGAC	1173
Db	1168	TGTCTCCAGGCTTGGCGCTGTGCTTTTGGCCCTATGAAGAAAGTTCTGCATCTCGAAGAC	1227
OY	1174	CTGTGGCGATCCATGAATTTCTGTGGCTACCGCCAGCTCATGATCTGAACAAGAACAGAG	1233
Db	1228	CTGTGGCGATCCATGAATTTCTGTGGCTTACCGCCAGCTCATGATCTGAACAAGAACAGAG	1287
OY	1234	AAACCTGTGCTGCATCTTCTGTCTACCTGSACTGTCCCTGAGCAAAATCTCAAGAGGG	1293

```

Db      1288  AAACCTGTGGATCTTCTGTCTACTGAGCTGTCCCTGAGCAATTCACCTAAGAGG 1347
QY      1294  TTGAGAAAGGCTTTTCTGCGACACCTGGTGTCAAGCCTCACTCACTTTCTCTGAATGCTTGTGA 1353
Db      1348  TTGAGAAAGGCTTTTCTGCGACACCTGGTGTCAAGCCTCACTCACTTTCTCTGAATGCTTGTGA 1407
QY      1354  GCCTTTGCCTACATCTCTGAGGCTTTAACTACATGTCTGTGGGTATCACATGAGTGTAGT 1413
Db      1408  GCCTTTGCCTACATCTCTGAGGCTTTAACTACATGTCTGTGGGTATCACATGAGTGTAGT 1467
QY      1414  TGTGTCCACAGTGTCTCAAGCAAGAGACTTTGTGTCCATGCTTGTGTCTAGAAAAAG 1473
Db      1468  TGTGTCCACAGTGTCTCAAGCAAGAGACTTTGTGTCCATGCTTGTGTCTAGAAAAAG 1527
QY      1474  ACTGGGGAAACCTTATGTGAGCAGCAATCCCAACAGTGAAGAAACAGGATTTGCTTCTTC 1533
Db      1528  ACTGGGGAAACCTTATGTGAGCAGCAATCCCAACAGTGAAGAAACAGGATTTGCTTCTTC 1587
QY      1534  TTTTCTGTGATCTTCTGTCTGTGGGACAGACTTGTAGAGACTTTGTGGCCTGTGAGGCTATTAA 1593
Db      1588  TTTTCTGTGATCTTCTGTCTGTGGGACAGACTTGTAGAGACTTTGTGGCCTGTGAGGCTATTAA 1647
QY      1594  GCAGAGACAGATATCAGTGAAGGAAATTGATCCATAAACCTCCCGTCCACATCTTGGCCAAATG 1653
Db      1648  GCAGAGACAGATATCAGTGAAGGAAATTGATCCATAAACCTCCCGTCCACATCTTGGCCAAATG 1707
QY      1654  GGGAAATGATCTTTCACCAAAAGAGCTCACAGACTTTTCCACAGAGATGCGAAATTCGAG 1713
Db      1708  GGGAAATGATCTTTCACCAAAAGAGCTCACAGACTTTTCCACAGAGATGCGAAATTCGAG 1767
QY      1714  CCCTTGGAAGTTCCCAATGGGATTCAGAGAGAAAGTGGGAACAGTTGGATGCTTACTT 1773
Db      1768  CCCTTGGAAGTTCCCAATGGGATTCAGAGAGAAAGTGGGAACAGTTGGATGCTTACTT 1826
QY      1774  ATGAGCTTGACCAT-ACAGCTATGCGTATGAGAAATGTGAACAAATCTCTG-ACAAA 1831
Db      1827  ATGAGCTTGACCATACAGCTATGCGTATGAGAAATGTGAACAAATCTCTGACAAA 1886
QY      1832  AGAGCAAGCTTAAATTACAGAGGTGGCTGGCTTGATTTGAAATATCA-TTTCCTTTTG 1890
Db      1887  AGAGCAAGCTTAAATTACAGAGGTGGCTGGCTGGCTTGATTTGAAATATCA-TTTCCTTTTG 1946
QY      1891  CATTTTCCCATTTACATAGAAAACTTTGACCTGTGAACCTTCCATCTGTAAATCTATAA 1950
Db      1947  CATTTTCCCATTTACATAGAAAACTTTGACCTGTGAAGC-TTCCATCTGTAAATCTATAA 2005
QY      1951  TTCCCAATTAAG 1962
Db      2006  TTCCCAATTAAG 2017

RESULT 3
US-09-190-911-2
; Sequence 2, Application US/09190911
; Patent No. 6365365
; GENERAL INFORMATION:
; APPLICANT: Bistrup, Annette
; APPLICANT: Rosen, Steven D.
; APPLICANT: Tangemann, Kirsten
; APPLICANT: Hemmerich, Stefan
; TITLE OF INVENTION: GLYCOSYL SULFOTRANSFERASE-3
; FILE REFERENCE: 6510-107CIP
; CURRENT APPLICATION NUMBER: US/09/190,911
; CURRENT FILING DATE: 1998-11-12
; EARLIER APPLICATION NUMBER: 09/045,284
; EARLIER FILING DATE: 1998-03-20
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 2032
; TYPE: DNA
; ORGANISM: H. sapiens

```


US-09-190-911-2

Query Match 87.4%; Score 1784.8; DB 4; Length 2032;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

QY 94 AGCCCAACCCAGAGCTCTTCCACTTACAGCACATAGTACTGCTTAAAAAATGAAGCT 153
 DB 148 AAGCCCGCTTGGAGAGTCTCCACTTCAAGCAATAGTACTGCTTAAAAAATGAAGCT 207
 QY 154 CCGGCTTTTGGTGTCCAGATGGCCACTGCTATCTTCCACATGACAGCA 213
 DB 208 CCGGCTTTTGGTGTCCAGATGGCCACTGCTATCTTCCACATGACAGCA 267
 QY 214 CAACATCAGCTCCCTGTATGAGGACAGCCGAGCCGATCAGCTGCTGTATGTC 273
 DB 268 CAACATCAGCTCCCTGTATGAGGACAGCCGAGCCGATCAGCTGCTGTATGTC 327
 QY 274 TTCCCTGGGCTGCTGCTCTTCTTTTGGGAGAGCTTTTGGGAGACCCAGATGTTT 333
 DB 328 TTCCCTGGGCTGCTGCTCTTCTTTTGGGAGAGCTTTTGGGAGACCCAGATGTTT 387
 QY 334 CTACCTGATGAGCCCGCTGGACGCTGAGATGACCTTCAAGCAGAGACCGCCGAT 393
 DB 388 CTACCTGATGAGCCCGCTGGACGCTGAGATGACCTTCAAGCAGAGACCGCCGAT 447
 QY 394 GCTGCACATGCTGTGCGGGATCTGATACGGCCGCTCTTCTTGTGACATGAGCTT 453
 DB 448 GCTGCACATGCTGTGCGGGATCTGATACGGCCGCTCTTGTGACATGAGCTT 507
 QY 454 TGAATGCTTACATGGAACCTGTGTCCCGGAGACAGTCCAGCTTTCAGTGGAGAACAG 513
 DB 508 TGAATGCTTACATGGAACCTGTGTCCCGGAGACAGTCCAGCTTTCAGTGGAGAACAG 567
 QY 514 CCGGGCCCTGTGTGCACTGCTGACATCATCCCAAGATGAATCATCCCGG 573
 DB 568 CCGGGCCCTGTGTGCACTGCTGACATCATCCCAAGATGAATCATCCCGG 627
 QY 574 GGTCTACGACAGCTCTGTGCACTGCAACAGCCCTTGTGAGTGGAGAACGCTCGG 633
 DB 628 GGTCTACGACAGCTCTGTGCACTGCAACAGCCCTTGTGAGTGGAGAACGCTCGG 687
 QY 634 CTCTCTACGACAGCTGTGCAAGAGAGGCGCTTCAACCTGACATCCCTCAACC 693
 DB 688 CTCTCTACGACAGCTGTGCAAGAGAGGCGCTTCAACCTGACATCCCTCAACC 747
 Y 694 GGTCTGAAAGACCCCTCCCTCAACCTGATATCTGCACTGTGTCGGGAGCCCGGGC 753
 DB 748 GGTCTGAAAGACCCCTCCCTCAACCTGATATCTGCACTGTGTCGGGAGCCCGGGC 807
 QY 754 CGTGTCCCTTCCCGAAGACGCAAAAGGAGATCTCATGATTGACAGTTCGATTGTAT 813
 DB 808 CGTGTCCCTTCCCGAAGACGCAAAAGGAGATCTCATGATTGACAGTTCGATTGTAT 867
 QY 814 GGGGCGAGCATGAGCAAAAACCTCAAGAGAGAGCAACCTACTATGATGAGTCAAT 873
 DB 868 GGGGCGAGCATGAGCAAAAACCTCAAGAGAGAGCAACCTACTATGATGAGTCAAT 927
 QY 874 CTGCCAAAGCAGCTGAGATCTACAAGACATCTCAGTCTTGCCTCAAGGCTCGAGGA 933
 DB 928 CTGCCAAAGCAGCTGAGATCTACAAGACATCTCAGTCTTGCCTCAAGGCTCGAGGA 987
 QY 934 AGCTTACCTGCTTGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCG 993
 DB 988 AGCTTACCTGCTTGTGCGCTATGAGAGACCTGGCTGAGCCCTGTGGCCAGACTTCCG 1047
 QY 994 AATGATGAATTCGCGGATGGAATTTCTGCCCTCTTCAAGCTGGTGGATGAATCAT 1053
 DB 1048 AATGATGAATTCGCGGATGGAATTTCTGCCCTCTTCAAGCTGGTGGATGAATCAT 1107
 QY 1054 CACCCGAGCAAGGCGATGGTGTGACACAGCTTTCACACAATATGCCAGGATGCCCTTAA 1113
 DB 1108 CACCCGAGCAAGGCGATGGTGTGACACAGCTTTCACACAATATGCCAGGATGCCCTTAA 1167

QY 1114 TGTCTCCAGAGCTTGGCGCTGCTTGTGCTTATGAAAAAGTTTCTGACTTCAGAAAGC 1173
 DB 1168 TGTCTCCAGAGCTTGGCGCTGCTTGTGCTTATGAAAAAGTTTCTGACTTCAGAAAGC 1227
 QY 1174 CTGTGGCGATGCCATTAATTTGCTGGGCTTACCGCCAGCTGACATCTGAAACAGACAGAG 1233
 DB 1228 CTGTGGCGATGCCATTAATTTGCTGGGCTTACCGCCAGCTGACATCTGAAACAGACAGAG 1287
 QY 1234 AAACCTGTGCTGATCTTGTCTTACCTGAGATGTCCTCGAGCAATCCACTAAGAGG 1293
 DB 1288 AAACCTGTGCTGATCTTGTCTTACCTGAGATGTCCTCGAGCAATCCACTAAGAGG 1347
 QY 1294 TTGAGAAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1353
 DB 1348 TTGAGAAGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1407
 QY 1354 GCTTCCCTTACATCTGTGAGCTTAACTTACATGCTGTGGGATTCACACTGAGTGTAGT 1413
 DB 1408 GCTTCCCTTACATCTGTGAGCTTAACTTACATGCTGTGGGATTCACACTGAGTGTAGT 1467
 QY 1414 TGTGTCACAGCTGCTCAAGCAGAGAGGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1473
 DB 1468 TGTGTCACAGCTGCTCAAGCAGAGAGGACTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1527
 QY 1474 ACTGGGAACTTATGTGAGCAGACATCCACAGATGAAACAGGCTATTTGCTCTTTC 1533
 DB 1528 ACTGGGAACTTATGTGAGCAGACATCCACAGATGAAACAGGCTATTTGCTCTTTC 1587
 QY 1534 TTTTCTTGTATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1593
 DB 1588 TTTTCTTGTATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1647
 QY 1594 GCAGCAGACAGTATCAGTGAATGTATCATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1653
 DB 1648 GCAGCAGACAGTATCAGTGAATGTATCATTAACCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1707
 QY 1654 GGGATGATCTTTCACCAAGAGCTCACACAGCTTTTCCACAGAGATGCCAATTCGTAG 1713
 DB 1708 GGGATGATCTTTCACCAAGAGCTCACACAGCTTTTCCACAGAGATGCCAATTCGTAG 1767
 QY 1714 CCTTGGAGTTCCCATGAGGATTCAGAGAGAGGAGGAGCAAGGTTGAGTCCACTT 1773
 DB 1768 CCTTGGAGTTCCCATGAGGATTCAGAGAGAGGAGGAGCAAGGTTGAGTCCACTT 1826
 QY 1774 ATGAGCTTGAACCAT-ACAGCTATCGGTAAATCAAGAAATATGAACAAATCTCTG-ACAAA 1831
 DB 1827 ATGAGCTTGAACCATCAAGCTATCGGTAAATCAAGAAATATGAACAAATCTCTGACAAA 1886
 QY 1832 AGAGCAAGCTCTTAAGTTCAGAGAGTGGCTGGCTTGAATTAATCA-TTTTCTTTG 1890
 DB 1887 AGAGCAAGCTCTTAAGTTCAGAGAGTGGCTGGCTTGAATTAATCACTTCCCTCTG 1946
 QY 1891 CATTTCCCATTAACATAGAAAACCTTGACCTGTGAACCTTGCAATGTTAATTAAGTAAAA 1950
 DB 1947 CATTTCCCATTAACATAGAAAACCTTGACCTGTGAACCTTGCAATGTTAATTAAGTAAAA 2005
 QY 1951 TTCCCAATTAAG 1962
 DB 2006 TTCCCAATTAAG 2017

RESULT 4
 US-08-899-514-1
 ; Sequence 1, Application US/08899514
 ; Patent No. 5910581
 ; GENERAL INFORMATION:
 ; APPLICANT: HABUCHI, OSAMI
 ; APPLICANT: FUKUTA, MASAKAZU
 ; TITLE OF INVENTION: POLYPEPTIDE OF GLYCOSAMINOGLYCAN
 ; TITLE OF INVENTION: SUFOTRANSFERASE ORIGINATING FROM HUMAN AND DNA CODING
 ; TITLE OF INVENTION: FOR THE POLYPEPTIDE
 ; NUMBER OF SEQUENCES: 9

```

CORRESPONDENCE ADDRESS:
ADDRESSEE: KNOBE, MARTENS, OLSON & BEAR, LLP
STREET: 620 NEWPORT CENTER DRIVE, SIXTEENTH FLOOR
CITY: NEWPORT BEACH
STATE: CALIFORNIA
COUNTRY: US
ZIP: 92660
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentlin
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/899, 514
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: DANIEL E. ALTMAN
REGISTRATION NUMBER: 34,115
REFERENCE/DOCKET NUMBER: FOYAM21.001AUS
TELECOMMUNICATION INFORMATION:
TELEPHONE: 714 760 0404
TELEFAX: 714 760 9502
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2156
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
ORIGINAL SOURCE:
ORGANISM: Human
TISSUE TYPE: Fetal brain
FEATURE:
NAME/KEY: CDS
LOCATION: 147..1583
IDENTIFICATION METHOD: S
US-08-899-514-1

Query Match      6.9%; Score 141.6; DB 2; Length 2156;
Best Local Similarity 50.5%; Pred. No. 8.2e-31;
Matches 488; Conservative 0; Mismatches 454; Indels 24; Gaps 5;

QY 257 CACGTGCTGTTCTGCTCTCTGCGCTCTGCTCTCTTTTGGGGCAGCTTTTGGG 316
DB 546 CACGTGCTGCTCATGCGCCACCGCGCTCTCTGCTGCGGAGCTTCTCAAC 605
QY 317 CAGACCCAGATGTTTCTTACCTGATGAGCCCGCGCTGCGACGTGTGA-----TGACC 370
DB 606 CAGCAGGGCACAATCTTCTTACCTCTTGCAGCGCTGTGACATCGACGACAGATGCC 665
QY 371 TTCAAGCAGAG-----CACCGCTGATGCTGCACATGCTGTGGGAGTGTATACGG 424
DB 666 TTGAGCGCGGGGGGCGCAAGCGCGGCTCGGCGCTGCTGATACGGAGCTGTCAAG 725
QY 425 GCCGCTTCTTGTGCGACATGAGCGCTTGTGATGCTCAATGAAAGCCTGTCCCGGAGA 484
DB 726 CAGCTTCTCTGCTGACACTGTAGTGTGAGACACTTCACAGCGCTCCCGAGAGAC 785
QY 485 CAGTCCAGCCTTCTTCACT-----GGGAGAACAGCGGGCGCTGTGCTGTGACCTGCC 538
DB 786 CACCTGACTCAGTTCTGATGCTCCGGGGCTCCAGCGCTCTCTGTGAGAGACCCCGTC 845
QY 539 TGTGACATCATCCACAAGATGAATATATCCCGGGGCTCAGTAGGCTCTGTGACGT 598
DB 846 TGTACGCCCTTCTGCAAGAGGTCTTGCAGAGATACACTCAAGAACCGCGCTGTGCGC 905
QY 599 CACAGCCCTTTTGAAGTGTGAGAGAGCGCTGCGCTCTTCAAGCAGCAGCTGTGCTCAAG 658
DB 906 CCCCTCAACGTGACGTGCGCGCAGAGGCGCTGCGCGCAAGAGACATGCGCCCTCAAG 965

```

```

QY 659 GAGGAGCGCTTCTTCAACCTGACAGTCCCTTACCGCGCTGTGAAGACCCCTCCCTAAC 718
DB 966 GCGGTGCCATCCGACACTGAGTCTTCTGACGCGCTGTGCGGAGACCCCGCTGAC 1025
QY 719 CTGATATCTGACACTGTGTCGGGACCCCGGCGCTGTTCGTTCCCGAAGCACA 778
DB 1026 CTGCGCGTCAATCCAGCTGTGCGGACCCCGGCGCTGTGCGCTGTGCGCTGTGCGC 1085
QY 779 AAGGAGATCTCATGATTTGACAGTGTGATGTGGGACGATGACCAAACTCAAG 838
DB 1086 TTCCCGCGCAAGTATTAAGACCTGGAAGAGTGGCTGACAGAGGCGGACAGCGCTG 1145
QY 839 AAGGAGACCAACCTACTATGTATGATGAGGTCTCCAAAGCAGCAGTGTAC 898
DB 1146 AGGGAAGAGA---GTCGACAGCGCTGTGGGCGCACTCGAGAGCATCCGCTGTCCCG 1202
QY 899 AAGACATCCAGTCTTGTCCCAAGGCGCTGTGAGAAAGCTACTGCTGTGCGCTTANGAG 958
DB 1203 GAGCTGGGGCTGCGGACCGCGCTGTGCTGGGGCGCTTACATGCTGTGCTGTACGAG 1262
QY 959 GACCTGTGCGAGCCCTGTGCGGCGGACCTTCCGAAATGTATGATTTGTGAGTGGAA 1018
DB 1263 GACGTGCGACGCGGCGCTGTGAGAAAGCGCGGAGATGTACCCGCTGTGCGGATGCC 1322
QY 1019 TTCTTGGCCCATCTTCAAGACCTGCGGTGATTAATCATCACCGAGGCAAGGCGATGGTAC 1078
DB 1323 CTGACCCCGCAGGTGTGAAGATGTGATCCAAAAGAACAGCAGG---CGGCCACAGCGC 1379
QY 1079 CACGCTTTCACACAAATGCGAGGATGCCCTTAATGTCTCCAGCTTGGCGCTGTCT 1138
DB 1380 AGCGCATCTACTCTCCAGCAGAAAGATCTCTGAGCAGTGTGAGAACTGCGGCTTACG 1439
QY 1139 TTGCGCTTGAAGAGTCTTCTGCACTTGAAGAGCTGTGCGAGTCCCATGATTTGCTG 1198
DB 1440 ATGCCCTTCAAGTGGCGCCAGAGTGTGAGGCGCCGCTGTGCGGCTGTGCTGTTC 1499
QY 1199 GCGTAC 1204
DB 1500 GCGTAC 1505

RESULT 5
US-09-015-188-1
Sequence 1, Application US/09015188C
Patent No. 639358
GENERAL INFORMATION:
APPLICANT: Williams, Kevin J
TITLE OF INVENTION: A Human Gene Encoding Human Chondrolectin
TITLE OF INVENTION: 6-Sulfotransferase.
FILE REFERENCE: JEFF-0231
CURRENT APPLICATION NUMBER: US/09/015.188C
CURRENT FILING DATE: 1998-01-29
NUMBER OF SEQ ID NOS: 17
SOFTWARE: Patentlin Ver. 2.0
SEQ ID NO 1
LENGTH: 2190
TYPE: DNA
ORGANISM: Homo sapiens
US-09-015-188-1

Query Match      6.6%; Score 135.2; DB 4; Length 2190;
Best Local Similarity 49.1%; Pred. No. 5.9e-29;
Matches 507; Conservative 0; Mismatches 483; Indels 42; Gaps 4;

QY 257 CACGCGCTGTTCTGCTCTCTGCGCTCTGCTCTCTTTTGGGGCAGCTTTTGGG 316
DB 308 CACATCTCATCTCTGCGCACCGCGCGCTCTCTCTCTGCTGTGCGCAGCTTTCAC 367
QY 317 CAGCACCAGATGTTTCTTCTTACCTGATGAGCCCGCGCTGACAGTGTGATGACCTTCAAG 376
DB 368 CAGACCTGAGACGTTCTTCTTACCTGTTTGAAGCCCTCTTCAACAGTCCAAACAGCTATC 427

```

```

OY 377 CAGAGACCGCCTGATGCTGACATAGCTGTGCG-----412
DB 428 CCCCCGCTTACCCAGGAGGAGCCGCGGAGCGGCTATGCTAGGCGGCACCGC 487
OY 413 GATCTGATACGGGCGCTCTTTGTCGATGAGCGCTTTGATCCCTACATGAACT 472
DB 488 GACCTCTCTGAGCGCTTACGACCTCTTACTCTCTGAGAACTACATCAAGCG 547
OY 473 GGTCCCGGAGACAGTCCAGCC-----TCTTCACTGGAGAGACCGGCGCTGTG 526
DB 548 CCGCGGTACACACACACACAGAGATCTCCGCGCGGCGGCGCGGCTCTCTG 607
OY 527 TGTGACCTGCTGTGACATCATCCACAGATGAATCATCC---CCGCGCTCACTGC 583
DB 608 TCCCGGCTGTGTGACACCTCCGCGGCGAGCGACCTGCTCTGAGAGGCGGACTGT 667
OY 584 AGGCTCTGTCGATCAACAGCCCTTGTGAGTGTGAGAGAGCCCTGCGCTCTACAG 643
DB 668 GTGCGGATGCGGCTACTCAACCTGACCGCGGCGGAGCGGTGCGGAGCGCAGC 727
OY 644 CAGTGTGCTCAAGAGGCTGCGCTTCTTCAACCTGACCTTACCCTGCTGTGAA 703
DB 728 CAGTGTGCTCAAGAGGCTGCGGCTGCGCGGAGTGAAGAGCTGCGCGCTGTGAA 787
OY 704 GACCGCTCCCTCAAGCTGATCTGTGACCTGTGCGGAGCCCGGCGCTGTCCGT 763
DB 788 GACCGCGATTAACCTCAAGGCTGATCAAGCTGTGCGGAGCCCGCGCTGTGCGT 847
OY 764 TCCCGAGACAGCAAGAGGAGATCTCATGATGACAGTGTGATGAGGCGAGCAT 823
DB 848 TCGCGAGAGAGACCTTCCGCGAGCTGACGCTGCGGCTGTGATACGCGACCGG 907
OY 824 GAGCAAAACTCAAGAGAGAGAGCAACCTTACTATGATGAGAGTCACTGTGCAAGC 883
DB 908 AGGAAACCTTAACTGAGAGCTGACGAGCTGACGAGCTGTGCGAGAGCTTCTCAAC 967
OY 884 CAGCTGAGATCTACAGACATCCAGTCTTGTGCGGAGCCCTGAGAGAGCTTACTG 943
DB 968 TCGGTCTTCAACGCGCTTACGCGGCGCGCGCTGCTCAAGGCGAAG-----TACATG 1018
OY 944 CTTGTGCTTATGAGAGCTGCTGAGCCCTGTGCGGAGCTTCCGATGTATGAA 1003
DB 1019 TTGTGTGCTTACGAGAGCTGTGCGAACCCTATGAGAGAGAGAGATCTACGG 1078
OY 1004 TTGTGTGAGTGAATTTTGTGCGGCTTGTGAGAGCTGTGATGAATCAACCGAGGC 1063
DB 1079 TTCTGTGCGATCCGCTGAGACAGCGCTGCGCGCTGTGATCAAGAAACAGCGGCG 1138
OY 1064 AAGGGCATGGTGACACGCTTTCACACAATGCCAGGGATGCCCTTATGTCTCCAG 1123
DB 1139 GACCCCTACCTGGGAGAGCAAAATAGGCGACCTGCGAAACTCGCGCGCACGCGCGAG 1198
OY 1124 GCTTGGCGCTGTCTTTGCCCTATGAAAGGTTTCTGCACTTCAAGAGCTGTGCGAT 1183
DB 1199 AAGTGGCGCTTCCGCTCTCTACGACATGTGCGCTTGTGCCAGAAACGCTGTCCAGAG 1258
OY 1184 GCGATGAATTTGTGCGGCTACCGCGACGTCAGATCTGAACAGAGAGAACTGTGTG 1243
DB 1259 GTGCTGTGCGGCTACAGATCTGAGATGCGCGCTGTGAGAGAGAGCTGAAGAACTCTG 1318
OY 1244 CTGATCTTCTG 1255
DB 1319 GTACAGCTGTG 1330

```

```

RESULT 6
US-08-655-878-1
; Sequence 1, Application US/08655878
; Patent No. 5827713
; GENERAL INFORMATION:
; APPLICANT: FUKUTA, MASARAZU
; APPLICANT: HABUCHI, OSAMI

```

```

? TITLE OF INVENTION: DNA CODING FOR SULFOTRANSFERASE
? NUMBER OF SEQUENCES: 9
? CORRESPONDENCE ADDRESS:
? ADDRESS:
? STREET:
? CITY:
? STATE:
? COUNTRY:
? ZIP:
? COMPUTER READABLE FORM:
? MEDIUM TYPE: Floppy disk
? COMPUTER: IBM PC compatible
? OPERATING SYSTEM: PC-DOS/MS-DOS
? SOFTWARE: Patent In
? CURRENT APPLICATION DATA:
? APPLICATION NUMBER: US/08/655,878
? FILING DATE:
? CLASSIFICATION: 435
? PRIOR APPLICATION DATA:
? APPLICATION NUMBER:
? FILING DATE:
? ATTORNEY/AGENT INFORMATION:
? NAME:
? REGISTRATION NUMBER:
? REFERENCE/DOCKET NUMBER:
? TELECOMMUNICATION INFORMATION:
? TELEPHONE:
? TELEFAX:
? INFORMATION FOR SEQ ID NO: 1:
? SEQUENCE CHARACTERISTICS:
? LENGTH: 2354
? TYPE: nucleic acid
? STRANDEDNESS: both
? TOPOLOGY: linear
? MOLECULE TYPE: cDNA to mRNA
? ORIGINAL SOURCE:
? ORGANISM: Chick
? TISSUE TYPE: Embryo chondrocyte
? FEATURE:
? NAME/KEY: CDS
? LOCATION: 211..1584
? IDENTIFICATION METHOD: P
? FEATURE:
? NAME/KEY: s1q-peptide
? LOCATION: 211..309
? IDENTIFICATION METHOD: P
? FEATURE:
? NAME/KEY: mat-peptide
? LOCATION: 310..1584
? IDENTIFICATION METHOD: P
? FEATURE:
? NAME/KEY: transmembrane domain
? LOCATION: 280..321
? IDENTIFICATION METHOD: P
? FEATURE:
? NAME/KEY: potential N-glycosylation site
? LOCATION: 394..402
? IDENTIFICATION METHOD: S
? FEATURE:
? NAME/KEY: potential N-glycosylation site
? LOCATION: 427..435
? IDENTIFICATION METHOD: S
? FEATURE:
? NAME/KEY: potential N-glycosylation site
? LOCATION: 493..501
? IDENTIFICATION METHOD: S
? FEATURE:
? NAME/KEY: potential N-glycosylation site
? LOCATION: 516..924
? IDENTIFICATION METHOD: S
? FEATURE:
? NAME/KEY: potential N-glycosylation site
? LOCATION: 1405..1413

```

```

; FEATURE: IDENTIFICATION METHOD: S
;
; FEATURE:
; NAME/KEY: potential N-glycosylation site
; FEATURE: LOCATION: 1537..1545
;
; FEATURE: IDENTIFICATION METHOD: S
US-08-655-878-1

```

Query Match	6.3%	Score 128.4	DB 1	Length 2354
Best Local Similarity	48.4%	Pred. No. 5.8e-27		
Matches 496; Conservative	0	Mismatches 501	Indels 27	Gaps 4

QY	250	GGCGATGCACAGTGGTGGTTCGCTTCCTCGGCGCTGGCCCTCTTTTGGGGGACACT	309
Db	543	GGCGGGGCAATCTCTGCTGATGGCCACACGCAACCGGCTCTCTTCGTTGGGGAGTT	602
QY	310	TTTTGGGACAGCCACCATGTTTCTTACCTGATGAGAGCCCGCTGGCAAGGTGATAC	369
Db	603	CTTCAACCAAGAGGGCAACATATTTCTACCTTTTGAGCCCTGGGCAATCGAAGAC	662
QY	370	CTTCA-----GCAGACACCGCTGGATGCTGCACATGGCTGTGGGAACT	417
Db	663	GGTACTTTTTGAGCCAGGGGGGGCAACGGCGTGGGCTCGGCGCTGTGATCCGAGCT	722
QY	418	GATACGGGCGCTCTTGTGCGACATGAGGGTCTTTGATGCTCATGAAAGCCGTGCC	477
Db	723	GGTCCACAGAGTCTCTCTCTGCGACCTCTACATCTGGAAGCTTCACTCACAGCGCC	782
QY	478	CCGGAGACAGTCCAG-----CTCTCTTCACTGGGAGAACAGCCGGGCGCTGTGTTGC	531
Db	783	CGAGGACACTTACTGCTCGCTCTGTTCCGGCGGGGCTCCAGCCACTACTCTGAGGA	842
QY	532	ACGTCGCTGTGACATCAATCCCAACAAAGTAATCATCCCCGGGGCTCACTCAGGCTCT	591
Db	843	GCCCCGTGCACACCCCAACCCCTCAAGAAAGTCTTTGAGAAATACCATGCAAGAACCGCG	902
QY	592	GTCGATCAACAGCCCTTTGAGTGTGGAGAAAGCCGTCGCCCTCTACAGCAAGTGGT	651
Db	903	CTGGGGGCTCTCAACATCAACGCGTGGAGCTGAAAGATCGCGGGGCAAGCAACATGGC	962
QY	652	GCTCAAGAGAGTGGCGCTTCTTCAACCTGACGTCCCTCTACCCGCTGCTGAAGACCCCTC	711
Db	963	CTTCAACACGGTGGGCAATCCGGGACGTGGAATTTCTGACGCCCTCGCGGAGAACCCCGC	1022
QY	712	CCTCAACCTGATATGTGACACCTGATGGCCGAGACCCCGGGCGGTTCCGTTCCCGAGA	771
Db	1023	GCTGGACCTCGGCTTTTTCACGCTGTGTGGGACCCACGATCCCGCTGGTCTC-----	1076
QY	772	ACGCACAAGGGAGATCTCATGATTGACAGTGCATGTGTGATGGGGCGATGAGCAAAA	831
Db	1077	GGCGATGTGGGCTTCTCGGGCAAGTACAGAGCTGGAAGAAATGGGGCGGCGAGGGGGA	1136
QY	832	ACTCAAGAGAGAACCAACCCCTACTATGTATGATGATGATCTGTGCCAAACCACTGGA	891
Db	1137	GGCCCCCTGTAGGAGAGACGAGGTGCAACGGCTCGGGGCAACTGCGAGACATCCGGCT	1198
QY	892	GATCTACAGACATCCAGTCTTGGCCCAAGGCCCTGCGAGAGAACGTACTGCTTGTGCG	951
Db	1197	GTCGGCGAGGTGGAGATCGGGGCAACCGCGGCTGGTGGAGGCGGTTTACATGCTGTGCG	1256
QY	952	CTATGAGAGACTGGCTGAGGCCCTGTGGCCCAAGACTTCCGAATGTATGAAATGTGCGG	1011
Db	1257	CTAGAGAGAGCTGGACAGGGGCGCGCTGCGCAAGCGCTGAGAGATGTACCCCTTCGCGCG	1318
QY	1012	ATTGGAATTTCTGGCCCATCTTTCAGACTTGGGTGCATTAATCAATCAACCCAGGCAAGGCAT	1071
Db	1317	CATCCACCCCAACGCGACAGGTGGAGAGATGATCCGGCCCAACAGCAGGCAACACAGGA	1378
QY	1072	GGGTGACCAAGCTTTCCACACAAATGCCAGGAGTCCCTTAATGTCTTCCAGGCTGGCG	1133
Db	1377	CAGCAAGAGGCTTTACTTCACAGCAG---AAGAACTCTCGAGAGAGTTTGAAGATGTGCG	1433
QY	1132	CTGGTCTTTGCCCTATGAAAAAGTTTTCAGACTTCAGAAACCTGTGGCATGTCATGCA	1191

D _b	1434	GTTTCAGCATCCCTTCAAGCGTGGGGACAGGTGGTGCAGAGACCCTGCGCAGGCCAGCCATGAG	1493
Q _y	1192	TTTTCCTGGGCTACCGCCACGTCATGATGAACAAGAACCCTGTTGGTAGTCT	1251
D _b	1494	GCTCTTGGCGCTACAAAGCTGGCGACAGTAGTCCAGAGAGCTGACCAAACCCCTGCTACAGCT	1553
Q _y	1252	TCTG 1255	
D _b	1554	GCTG 1557	

RESULT 7
US-09-263-023-3
; Sequence 3, Application US/09263023

```

1 GENERAL INFORMATION:
2 APPLICANT: Uchimura, Kenji
3 APPLICANT: Muramatsu, Hideki
4 APPLICANT: Kadomatsu, Kenji
5 APPLICANT: Kanmaji, Keiji
6 APPLICANT: Habuchi, Osami
7 APPLICANT: Muramatsu, Takashi
8 TITLE OF INVENTION: POLYPEPTIDE OF N-ACETYLGLUCOSAMINE-6-0-SULFOTRANSFERASE AND
9 TITLE OF INVENTION: DNA ENCODING THE SAME
10 FILE REFERENCE: TOYAMA1.001AUS
11 CURRENT APPLICATION NUMBER: US/09/263,023
12 CURRENT FILING DATE: 1999-03-05
13 EARLIER APPLICATION NUMBER: JP 10-54007
14 EARLIER FILING DATE: 1998-03-05
15 EARLIER APPLICATION NUMBER: JP 10-177844
16 EARLIER FILING DATE: 1998-06-24
17 NUMBER OF SEQ ID NOS: 10
18 SOFTWARE: FastSeq for Windows Version 3.0

```

```

; LENGTH: 2409
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (390)...(1841)
; US-09-263-023-3

```

Query Match	5.8%	Score 119.4	DB 3	Length 2409
Best Local Similarity	49.1%	Pred. No. 2.4e-24		
Matches 508; Conservative	0	Mismatches 446;	Indels 81;	Gaps 4;

QY	261	TGCGGATTCGTCCTTCCGAGCGCTGGCTCTCTTTGTGTGGGACGTTTTTGGGACG	320
Db	751	TGTACGTTCACCAACGTCGCGCTCTGGCTGTGTGTTTGTGGGACGTTATTCAACG	810
QY	321	ACCAGATGTTTCTTACCTGATGAGCGCCGCGTCGACGTGTGATGACCTTCAACG	380
Db	811	ATCCGAGGCTCTTCTCTTCTACGACGACGATGTGGCATGTATGGCAAAACGTATCCG	870
QY	381	GCACCGCTTGGATCTGCATGCGCTGTGCGGATCTATACGGGCGCTCTTGTGTGCG	440
Db	871	GCGACGCGGTTTCCCTGCGAGGGGACGCGCGGACATCTAGCGCTTTTACCGCTACG	930
QY	441	ACATGAGCGCTTTTGATGATCCATCAATGGAACCTGGGCCCCGGAGAC-----AGT	488
Db	931	ACCTCTCTGCTTCTTCCAGTTGATATAGCCCCGCGGCGAGCGGGGCGACACTCAACG	990
QY	489	CCAGCCTCTTTCAGTGGGAGAACGCGGGCCCTGTGTTTGTGCACCTGCTGTGATCA	548
Db	991	TGGCATCTTCGGGCGACGACCAACAAAGTGTGTGCTGTGCACCACTGATGCCCGCT	1050
QY	549	TCCCAACAAGATGAATCATCCCCGGGCTCACTGAGGCTC---CTGTGCAGTCAACAGC	605
Db	1051	ACCCCAAGGAGTCTGGGATTGTGTGCGACGCCGCTGTGCAGAAAGTGCCTGCCACG	1110
QY	606	CCTTTGAAGTGGTGAAGAGCGCTCCGCGCTCTACAGCAGCAGTGGTGCTCAAGAGTGC	665
Db	1111	GCTGGGCGCTTTTGAGAGGAGATGCGCGCAAGTACCGCACTACTATCAATAAAGGTGTGC	1170

QY 535 TGCCTGTGATCATTCACCAAGAGAAATCATCCCGGGCTCACTGCAGGCTCC--T 591
Db 1114 ACTGTCTCTGCTGCTACCGCAAGAGAGTCTCGGACTGTGGAGCAGACCGGTGTGCAAAA 1173
QY 592 GTGCAGTCAACAGCCCTTTAGAGTGTGTGAGAGAGGCTCCGCTCTACAGCCACCTGGT 651
Db 1174 GTGCCACCTCAACAGCTGTGACGCTTCGAGAGAGGTGTGCAAGTACCGCAGGTGTGT 1233
QY 652 GCTCAAGAGAGTGGCTTCTTCAACCTGCAGTCCCTCAACCGCTGCTCAAGACCCCTC 711
Db 1234 TATCAAGGCGGTGGGGGTCTTGATGTGGCTGTGTGGGGCGGCTCTTAAGATCTCAG 1293
QY 712 CCTCAACCTGCAATTCGTGCTGCTGCTCGGAGACCGGCGCTTCCGTTCCCGAGA 771
Db 1294 CTTGACACTCAAGGCTATCATCTACTGATCTGCTGCTGCTGCTGCCAGCTCCGCAAT 1353
QY 772 AGCAGCAAGAGAGATTCATGATGATGATGATGATGATGATGATGATGATGATGATG 831
Db 1354 CCGCTCGCTCAACGCGCTCATCCGGGAAAGCTTACAGGTGTGTGCGAAGCGGGAGCAAG 1413
QY 832 ACTCAA 837
Db 1414 AGCCCA 1419

RESULT 11

US-09-040-984-63
Sequence 63, Application US/09040984
Patent No. 6210883
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Wang, Tonglong
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR DIAGNOSIS
TITLE OF INVENTION: OF LUNG CANCER
NUMBER OF SEQUENCES: 86
CORRESPONDENCE ADDRESS:
ADDRESSEE: SEED and BERRY LLP
STREET: 6300 Columbia Center, 701 Fifth Avenue
CITY: Seattle
STATE: WA
COUNTRY: USA
ZIP: 98104
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: DOS
SOFTWARE: FASTSEQ for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/040,984
FILING DATE: 18-MAR-1998
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Maki, David J.
REGISTRATION NUMBER: 31,392
REFERENCE NUMBER: 210121.456
TELECOMMUNICATION INFORMATION:
TELEPHONE: 206-622-4900
TELEFAX: 206-282-6031
TELEX:
INFORMATION FOR SEQ ID NO: 63:
SEQUENCE CHARACTERISTICS:
LENGTH: 731 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
US-09-040-984-63

Query Match 3.1%; Score 62.6; DB 3; Length 731;
Best Local Similarity 58.2%; Pred. No. 3.7e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;
648 TGGTCTCAAGAGAGTGGCGCTTCTTCACCTGCAGTCCCTCTACCGCTGTGAGAAAGACC 707

Db 3 TACTCATAAAGGATGGTGGCGGTCTTCAGCTGGCGGTCTTGGCGCACTGCTCGAGACC 62
QY 708 CTTCCCTCAACGATGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 767
Db 63 CGGCGCTGAGACCTCAAGGATCATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 122
QY 768 GAGACGCAACAAAGGAGATTCATGATGATGATGATGATGATGATGATGATGATGATG 827
Db 123 GATCCGCTGCGCGCCACAGGCTCATCTGAGAGAGCTTACAGGTGTGTGCGAGCGGAGACC 182
QY 828 AAAACTCA 836
Db 183 GCGAGCTCA 191

RESULT 12

US-09-123-912-63
Sequence 63, Application US/09123912A
Patent No. 6312695
GENERAL INFORMATION:
APPLICANT: Reed, Steven G.
APPLICANT: Wang, Tonglong
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY OF LUNG CANCER
FILE REFERENCE: 210121.455C1
CURRENT APPLICATION NUMBER: US/09/123,912A
CURRENT FILING DATE: 1998-07-27
PRIOR APPLICATION NUMBER: 09/040,802
PRIOR FILING DATE: 1998-03-18
NUMBER OF SEQ ID NOS: 114
SOFTWARE: Patentln Ver. 2.0
SEQ ID NO 63
LENGTH: 731
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: modified_base
LOCATION: (236)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (249)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (263)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (288)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (312)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (317)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (323)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (326)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (337)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (352)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (362)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base
LOCATION: (370)
OTHER INFORMATION: Where n is a, c, g or t
NAME/KEY: modified_base

1	OTHER INFORMATION: Where n is a, c, g or t
2	NAME/KEY: modified_base
3	LOCATION: (615)
4	OTHER INFORMATION: Where n is a, c, g or t
5	NAME/KEY: modified_base
6	LOCATION: (627)
7	OTHER INFORMATION: Where n is a, c, g or t
8	NAME/KEY: modified_base
9	LOCATION: (632)
10	OTHER INFORMATION: Where n is a, c, g or t
11	NAME/KEY: modified_base
12	LOCATION: (640)
13	OTHER INFORMATION: Where n is a, c, g or t
14	NAME/KEY: modified_base
15	LOCATION: (641)
16	OTHER INFORMATION: Where n is a, c, g or t
17	NAME/KEY: modified_base
18	LOCATION: (644)
19	OTHER INFORMATION: Where n is a, c, g or t
20	NAME/KEY: modified_base
21	LOCATION: (654)
22	OTHER INFORMATION: Where n is a, c, g or t
23	NAME/KEY: modified_base
24	LOCATION: (660)
25	OTHER INFORMATION: Where n is a, c, g or t
26	NAME/KEY: modified_base
27	LOCATION: (663)
28	OTHER INFORMATION: Where n is a, c, g or t
29	NAME/KEY: modified_base
30	LOCATION: (665)
31	OTHER INFORMATION: Where n is a, c, g or t
32	NAME/KEY: modified_base
33	LOCATION: (671)
34	OTHER INFORMATION: Where n is a, c, g or t
35	NAME/KEY: modified_base
36	LOCATION: (678)
37	OTHER INFORMATION: Where n is a, c, g or t
38	NAME/KEY: modified_base
39	LOCATION: (692)
40	OTHER INFORMATION: Where n is a, c, g or t
41	NAME/KEY: modified_base
42	LOCATION: (697)
43	OTHER INFORMATION: Where n is a, c, g or t
44	NAME/KEY: modified_base
45	LOCATION: (698)
46	OTHER INFORMATION: Where n is a, c, g or t
47	NAME/KEY: modified_base
48	LOCATION: (699)
49	OTHER INFORMATION: Where n is a, c, g or t
50	NAME/KEY: modified_base
51	LOCATION: (704)
52	OTHER INFORMATION: Where n is a, c, g or t
53	NAME/KEY: modified_base
54	LOCATION: (705)
55	OTHER INFORMATION: Where n is a, c, g or t
56	NAME/KEY: modified_base
57	LOCATION: (712)
58	OTHER INFORMATION: Where n is a, c, g or t
59	NAME/KEY: modified_base
60	LOCATION: (714)
61	OTHER INFORMATION: Where n is a, c, g or t
62	NAME/KEY: modified_base
63	LOCATION: (717)
64	OTHER INFORMATION: Where n is a, c, g or t
65	NAME/KEY: modified_base
66	LOCATION: (718)
67	OTHER INFORMATION: Where n is a, c, g or t
68	NAME/KEY: modified_base
69	LOCATION: (719)
70	OTHER INFORMATION: Where n is a, c, g or t
71	NAME/KEY: modified_base
72	LOCATION: (723)

Query Match 3.1%; Score 62.6; DB 4; Length 731;
Best Local Similarity 58.2%; Pred. No. 3.7e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 648 TGGTGTCTCAAGGAGGTGGCTTCTTCAACCTGACGTCCTTACCCGCTGCTGAAGAGCC 707
DB 3 TAGTCATAAAGGTGTGCGCGCTTCTGACGTGGCGCTTGGCGCAGCTGCTCGAGACC 62
QY 708 CCTCCCTCAACCTGCATATCTGACACCTGTCGGGAGCCCGGGCGCTGTTCCGTTCCC 767
DB 63 CGGCCCTGACCTCAAGGTCAATCCTGCTGATCCCGCGCGGTGGGAGATTAC 122
QY 768 GAGAAGCGCAAAAGGAGATCTCATGATTCAGACATGTCGATTTGATGGGCGACATGAGC 827
DB 123 GGATCGGCTCGGCGCCACGCTCATCTCCGTGAGAGCCTACAGGTGTGGCGCAGCCGAGACC 182
QY 828 AAAAAGCTCA 836
DB 183 GCGAGCTCA 191

RESULT 13
US-09-643-597-63
Sequence 63, Application US/09643597
Patent No. 6426072
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
APPLICANT: Li, Samuel X.
APPLICANT: Wang, Aijun
APPLICANT: Skelky, Yasir A.W.
APPLICANT: Henderson, Robert A.
APPLICANT: McNeill, Patricia D.
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR THE THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.455C11
CURRENT APPLICATION NUMBER: US/09/643,597
CURRENT FILING DATE: 2000-08-21
NUMBER OF SEQ ID NOS: 369
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 63
LENGTH: 731
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(731)
OTHER INFORMATION: n = A,T,C or G
US-09-643-597-63

Query Match 3.1%; Score 62.6; DB 4; Length 731;
Best Local Similarity 58.2%; Pred. No. 3.7e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 648 TGGTGTCTCAAGGAGGTGGCTTCTTCAACCTGACCTCCCTTACCCGCTGCTGAAGAGCC 707
DB 3 TAGTCATAAAGGTGTGCGCGCTTCTGACGTGGCGCTTGGCGCAGCTGCTCGAGACC 62
QY 708 CCTCCCTCAACCTGCATATCTGACACCTGTCGGGAGCCCGGGCGCTGTTCCGTTCCC 767
DB 63 CGGCCCTGACCTCAAGGTCAATCCTGCTGATCCCGCGCGGTGGGAGATTAC 122
QY 768 GAGAAGCGCAAAAGGAGATCTCATGATTCAGACATGTCGATTTGATGGGCGACATGAGC 827
DB 123 GGATCGGCTCGGCGCCACGCTCATCTCCGTGAGAGCCTACAGGTGTGGCGCAGCCGAGACC 182
QY 828 AAAAAGCTCA 836
DB 183 GCGAGCTCA 191

RESULT 14
US-09-480-884A-63
Sequence 63, Application US/09480884A
Patent No. 6482597
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Hosken, Nancy
APPLICANT: Fanger, Gary R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.455C6
CURRENT APPLICATION NUMBER: US/09/480,884A
CURRENT FILING DATE: 2001-08-27
NUMBER OF SEQ ID NOS: 330
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 63
LENGTH: 731
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)...(731)
OTHER INFORMATION: n = A,T,C or G
US-09-480-884A-63

Query Match 3.1%; Score 62.6; DB 4; Length 731;
Best Local Similarity 58.2%; Pred. No. 3.7e-08;
Matches 110; Conservative 0; Mismatches 79; Indels 0; Gaps 0;

QY 648 TGGTGTCTCAAGGAGGTGGCTTCTTCAACCTGACGTCCTTACCCGCTGCTGAAGAGCC 707
DB 3 TAGTCATAAAGGTGTGCGCGCTTCTGACGTGGCGCTTGGCGCAGCTGCTCGAGACC 62
QY 708 CCTCCCTCAACCTGCATATCTGACACCTGTCGGGAGCCCGGGCGCTGTTCCGTTCCC 767
DB 63 CGGCCCTGACCTCAAGGTCAATCCTGCTGATCCCGCGCGGTGGGAGATTAC 122
QY 768 GAGAAGCGCAAAAGGAGATCTCATGATTCAGACATGTCGATTTGATGGGCGACATGAGC 827
DB 123 GGATCGGCTCGGCGCCACGCTCATCTCCGTGAGAGCCTACAGGTGTGGCGCAGCCGAGACC 182
QY 828 AAAAAGCTCA 836
DB 183 GCGAGCTCA 191

RESULT 15
US-09-542-615A-63
Sequence 63, Application US/09542615A
Patent No. 6518256
GENERAL INFORMATION:
APPLICANT: Wang, Tongtong
APPLICANT: Fan, Liqun
APPLICANT: Kalos, Michael D.
APPLICANT: Bangur, Chaitanya S.
APPLICANT: Hosken, Nancy A.
APPLICANT: Fanger, Gary R.
TITLE OF INVENTION: COMPOUNDS AND METHODS FOR THERAPY
TITLE OF INVENTION: AND DIAGNOSIS OF LUNG CANCER
FILE REFERENCE: 210121.455C8
CURRENT APPLICATION NUMBER: US/09/542,615A
CURRENT FILING DATE: 2000-04-14
NUMBER OF SEQ ID NOS: 350
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 63
LENGTH: 731
TYPE: DNA
ORGANISM: Homo sapien
FEATURE:

THIS PAGE BLANK (USPTO)

GenCore version 5.1.6
Copyright (c) 1993 - 2003 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 11:17:45 ; Search time 3864.66 Seconds
(without alignments)
12848.224 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 2043
Sequence: 1 gaattccatgctgttggtga.....tggatcccaaaaaaaaaa 2043

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 22781392 seqs, 1215238056 residues

Total number of hits satisfying chosen parameters: 45562784

Index DB seq length: 0
aximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*
1: em_estba:*
2: em_esthum:*
3: em_estln:*
4: em_estnu:*
5: em_estov:*
6: em_estpl:*
7: em_estro:*
8: em_estc:*
9: gb_estl:*
10: gb_estc:*
11: gb_estc3:*
12: gb_estc4:*
13: gb_estc5:*
14: gb_estc6:*
15: em_estfun:*
16: em_estom:*
17: em_gss_hum:*
18: em_gss_inv:*
19: em_gss_pln:*
20: em_gss_vtl:*
21: em_gss_fun:*
22: em_gss_mam:*
23: em_gss_mus:*
24: em_gss_pro:*
25: em_gss_rtd:*
26: em_gss_phg:*
27: em_gss_vrl:*
28: gb_gssl:*
29: gb_gss2:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	ID	Description
1	815.6	39.9	954	12 B1823850 603039012
2	760.6	37.2	1923	11 AK009113 Mus muscu
3	647.6	31.7	668	9 AL709927 DNF2P860
4	581.4	28.5	583	2 HSM082755 Bx490456 Homo sap1

Result No.	Score	Query Length	ID	Description
5	577	28.2	669	13 B0684395 UI-CF-ENO
6	507	24.8	593	12 B0692922 B0692922
7	499.8	24.5	571	12 B0692922 B0692922
8	496.2	24.3	680	10 B0692922 B0692922
9	476	23.3	553	12 B0692922 B0692922
10	476	23.3	553	12 B0692922 B0692922
11	385.2	18.9	783	12 B0692922 B0692922
12	368.6	18.0	695	12 B0692922 B0692922
13	364.4	17.8	852	12 B0692922 B0692922
14	363.4	17.8	955	10 B0692922 B0692922
15	337.8	16.5	419	9 B0692922 B0692922
16	325.6	15.9	902	13 B0692922 B0692922
17	288	14.1	362	9 B0692922 B0692922
18	282.6	13.8	834	13 B0692922 B0692922
19	274.4	13.4	571	12 B0692922 B0692922
20	274.4	13.4	695	10 B0692922 B0692922
21	271.2	13.3	358	10 B0692922 B0692922
22	263.6	12.9	735	10 B0692922 B0692922
23	253.8	12.4	692	14 B0692922 B0692922
24	252.4	12.4	620	9 B0692922 B0692922
25	239.2	11.7	507	12 B0692922 B0692922
26	235	11.6	804	13 B0692922 B0692922
27	235	11.5	1067	29 B0692922 B0692922
28	221	10.8	525	10 B0692922 B0692922
29	217.8	10.7	904	13 B0692922 B0692922
30	216.8	10.6	773	13 B0692922 B0692922
31	213.8	10.5	768	13 B0692922 B0692922
32	209.2	10.2	703	14 B0692922 B0692922
33	209	10.2	813	14 B0692922 B0692922
34	200.2	9.8	407	9 B0692922 B0692922
35	192.6	9.4	849	29 B0692922 B0692922
36	192.4	9.4	692	13 B0692922 B0692922
37	192.2	9.4	687	14 B0692922 B0692922
38	167	8.2	568	9 B0692922 B0692922
39	158.8	7.8	657	29 B0692922 B0692922
40	156.6	7.7	555	10 B0692922 B0692922
41	156.2	7.6	517	10 B0692922 B0692922
42	156.2	7.6	656	9 B0692922 B0692922
43	155.2	7.6	433	28 B0692922 B0692922
44	154.4	7.6	543	13 B0692922 B0692922
45	147.2	7.2	380	12 B0692922 B0692922

ALIGNMENTS

RESULT 1
B1823850
LOCUS
DEFINITION
603039012F1 NIH_MGC_115 Homo sapiens cDNA clone IMAGE:5179826 5',
mRNA sequence.
ACCESSION
B1823850
VERSION
B1823850.1 GI:15935400
KEYWORDS
EST.
SOURCE
Homo sapiens (human)
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
NIH-MGC http://mgc.ncl.nih.gov/
TITLE
National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL
Unpublished
COMMENT
Contact: Robert Strausberg, Ph.D.
Email: cgaps-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LNL)
DNA Sequencing by: Incyte Genetics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LNL at:
http://image.llnl.gov
Plate: L1448 row: h column: 03
High quality sequence stop: 856.

FEATURES
source

Location/Qualifiers
1. .954
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="IMAGE:5179826"
/lab_host="DH10B"
/clone_id="NH_MGC_115"
/note="Organ: pooled brain, lung, testis; Vector: PCMV-SPORT6; Site: 1: NotI; Site 2: EcoRV (destroyed); RNA source anonymous pool of 6 male brains, age range 23-27; 1 male lung, age 27; and 1 male testis, age 69. Library is oligo-dT primed and directionally cloned (EcoRV site is destroyed upon cloning). Average insert size 1.8 kb, insert size range 1-3 kb. Library is normalized and enriched for full-length clones and was constructed by C. Gruber (Invitrogen). Research Genetics tracking code 021. Note: this is a NIH_MGC Library."
BASE COUNT 199 a 297 c 250 g 208 t
ORIGIN

Query Match 39.9%; Score 815.6; DB 12; Length 954;
Best Local Similarity 97.8%; Pred. No. 8.1e-205;
Matches 933; Conservative 0; Mismatches 9; Indels 12; Gaps 10;

51 GCGCATGGCCGGCTAGCAGTACGCTCTCAAAAGCAGGAGGAGCCCAAGCCACAGG 110
1 GCGCATGGCCGGCTAGCAGTACGCTCTCAAAAGCAGGAGGAGCCCAAGCCACAGG 60
111 TCTTCCACTTTCAGCAATAGTACTGCTTAAATGAGTCCGCTGTTTGGTTT 170
61 TCTTCCACTTTCAGCAATAGTACTGCTTAAATGAGTCCGCTGTTTGGTTT 120
171 CCCAGATGGCCATCTGCTTATCTTCCACATGATACGCCACAACATCAGCTCCCT 230
121 CCCAGATGGCCATCTGCTTATCTTCCACATGATACGCCACAACATCAGCTCCCT 180
231 CATATAAGCAGACGCCAGGCGCATGACAGTGTCTGTTCTTCTGCGCTGCT 290
181 CATATAAGCAGACGCCAGGCGCATGACAGTGTCTGTTCTTCTGCGCTGCT 240
291 CTTCTTTTGGGGAGAGCTTTTGGGAGCAGCCAGATGTTTCTACCTGATGAGAGCCG 350
241 CTTCTTTTGGGGAGAGCTTTTGGGAGCAGCCAGATGTTTCTACCTGATGAGAGCCG 300
351 CCTGCGACGTGTGATGACCTTCAAGCAGAGCAGCCGCTGATGCTGCATGGCTGTC 410
301 CCTGCGACGTGTGATGACCTTCAAGCAGAGCAGCCGCTGATGCTGCATGGCTGTC 360
411 GGGATCTGATAGGGCCCTCTCTTGTGTGACATGAGCTCTTGTATGCTCATGGAAC 470
361 GGGATCTGATAGGGCCCTCTCTTGTGTGACATGAGCTCTTGTATGCTCATGGAAC 420
471 CTGTGTCGCCGAGACAGTCCAGCTCTTCACTGGGAGAACAGCGGCGCTGTGTTG 530
421 CTGTGTCGCCGAGACAGTCCAGCTCTTCACTGGGAGAACAGCGGCGCTGTGTTG 480
531 CACCTGCTGTGATCATCCACAAGATGAATATATCCCGGGGCTCACTGACAGCTCC 590
481 CACCTGCTGTGATCATCCACAAGATGAATATATCCCGGGGCTCACTGACAGCTCC 540
591 TGTGAGTCAACAGCCCTTGAAGTGTGGAGAGGCGCTGCTCTCTACACCAAGCTG 650
541 TGTGAGTCAACAGCCCTTGAAGTGTGGAGAGGCGCTGCTCTCTACACCAAGCTG 600
651 TGTGAGTCAACAGCCCTTGAAGTGTGGAGAGGCGCTGCTCTCTACACCAAGCTT 710
601 TGTGAGTCAACAGCCCTTGAAGTGTGGAGAGGCGCTGCTCTCTACACCAAGCTT 660
711 CCTCAACCTGCATATCTGACAC--TGTGTCGGGAGACCCCGGCGCTGTTCGTTCCGA 769
661 CCTCAACCTGCATATCTGACACCTTGTGTCGGGAGACCCCGGCGCTGTTCGTTCCGA 720

QY 770 GAACGCA--CAAGGAGATCTCATGATTGA--CAGTCGATTTGATGGGCGACATGAC 827
DB 721 GAAGCAGCAACCAAGGAGATCTCATGATTGATGACAGTGTGATGGGCGACATGAC 780
QY 828 AAAAA--CTCAGAAAGGAGGAGAC--AACCTACTATGTATG--CAGTCACTTCCCAA-- 881
DB 781 AAAAACTCAGAAAGGAGGAGACCAAACTTACTATGTATGTCAGAGTCACTTCCCAAAG 840
QY 882 GCGAGCTGAGATCTAC--NAGACATCCAGTCTTGCCCAAGGCGCTGAGGAGAGCTAC 940
DB 841 CCAGCTGAGATCTTCAAAAGACATCCAGTCTTGCCCAAGGCGCTGAGGAGAGCTAC 900
QY 941 CTGCTTGTGCG--CTATGAGGACCTGCTTGA--GCCCTGTGCGCCAGACTTCC 992
DB 901 CTGCTTGTGCGCAATGAGGAGCTGCTGAGGAGGCGCTGCTGCGCCAGACTTCC 954

RESULT 2
AK009113
LOCUS

DEFINITION AK009113 1923 bp mRNA linear HTC 05-DEC-2002
Mus musculus adult male tongue cDNA, RIKEN full-length enriched library, clone:2310003G18 product:carbohydrate (chondroitin 6-sulfate) sulfotransferase 4, full insert sequence.

ACCESSION AK009113.1 GI:12843701
VERSION AK009113.1
KEYWORDS HTC; CAP trapper.
SOURCE Mus musculus (house mouse)
ORGANISM Mus musculus

REFERENCE

1 Carninci, P., and Hayashizaki, Y.
High-efficiency full-length cDNA cloning
Meth. Enzymol. 303, 19-44 (1999)
99279253
10349636

REFERENCE

2 Carninci, P., Shibata, Y., Hayatsu, N., Sugahara, Y., Shibata, K., Itoh, M., Konno, H., Okazaki, Y., Muramatsu, M., and Hayashizaki, Y.
Normalization and subtraction of cap-trapper-selected cDNAs to prepare full-length cDNA libraries for rapid discovery of new genes
Genome Res. 10 (10), 1617-1630 (2000)
20493374
11042159

REFERENCE

3 Shibata, K., Itoh, M., Alzawa, K., Nagaoka, S., Sasaki, N., Carninci, P., Konno, H., Akiyama, J., Nishi, K., Kitsuana, T., Tashiro, H., Itoh, M., Sumi, N., Ishii, Y., Nakamura, S., Hazama, M., Nishino, T., Harada, A., Yamamoto, R., Matsumoto, H., Sakaguchi, S., Ikegami, T., Kashiwagi, K., Fujiwara, S., Inoue, K., Togawa, Y., Izawa, M., Ohara, E., Watahiki, M., Yoneda, Y., Ishikawa, T., Ozawa, K., Tanaka, T., Matsura, S., Kawai, J., Okazaki, Y., Muramatsu, M., Inoue, Y., Kira, A., and Hayashizaki, Y.
RIKEN integrated sequence analysis (RISA) system--384-format sequencing pipeline with 384 multichannel sequencer
Genome Res. 10 (11), 1757-1771 (2000)
20530913
11078661

JOURNAL MEDLINE PUBMED

REFERENCE

4 Kawai, J., Shinagawa, A., Shibata, K., Yoshino, M., Itoh, M., Ishii, Y., Arakawa, T., Hara, A., Fukunishi, Y., Konno, H., Adachi, J., Fukuda, S., Alzawa, K., Izawa, M., Nishi, K., Kiyosawa, H., Kondo, S., Yamanka, I., Saito, T., Okazaki, Y., Gojobori, T., Bono, H., Kasukawa, T., Saito, R., Kadota, K., Matsuda, H., Ashburner, M., Batalov, S., Casavant, T., Fleischmann, W., Gaasterland, T., Glast, C., King, B., Kochiwa, H., Kuehl, P., Lewis, S., Matsuo, Y., Nishino, I., Pesole, G., Quackenbush, J., Schriml, L. M., Stabli, F., Suzuki, R., Tomita, M., Wagner, L., Washio, T., Sakai, K., Okido, T., Furuno, M., Aono, H., Baldarelli, R., Barsh, G., Blake, J., Boffelli, D., Bojunga, A., Carninci, P., de Bonaldo, M. F., Brownstein, M. J., Bull, C., Fletcher, C., Fujita, M., Gariboldi, M., Gustinich, S., Hill, D., Hofmann, M., Hume, D. A., Kamiya, M., Lee, N. H., Lyons, P., Marchionni, L., Mashima, J., Mazzarelli, J., Mombert, P., Nori, P., Ring, B., Ringwald, M., Rodriguez, I., Sakamoto, N., Sasaki, H.,


```

Db      981 GATTTTGGCCACCCTCCAAACAGGGTTTACATGTACCCCGGCAAGGGCAGTGGT 1040
OY      1076 GACCAAGCTTTTCCACAAATGCGAGGATGCGCTTATGTCTCCAGGCTTGGGCTGG 1135
Db      1041 CAGCATGCTTTCCATCTAACGCCAGAACGCCCTCAAGCTCTCTAGGCGTGGCTTGG 1100
OY      1136 TCTTGGCTTATGAAAGGTTTCTGACTTCAGAAAGCTGTGGCGATCCATGAATTTG 1195
Db      1101 TCTTACCTTACGAAAGGTTTCCAGCTTCAAGATGCTGGCGGTAGCTATGATTTG 1160
OY      1196 CTGGGCTACGGCAGCTGATCTGAAACAAGAGAAACCTGTGTGATCTTCG 1255
Db      1161 CTGGGATACCTCCAGCTCAGATCTCAACAAGAACAGCACTGTCTGGATCTTGG 1220
OY      1256 TCTACCTGAGCTGTCCCTGAGCAATCACTAAGAGGTTGAGAGGCTTGTGCTG-CCA 1313
Db      1221 TCTCTCCCTCATATCTTGGGAGGCTTCCGAGAAAGTTAAGAGGCTGTCTGACCC 1280
OY      1314 CCTGGGTAGGCTCAGTCACTTCTCTAATGCTTCTAGCCCTGCTACATCTGTAG 1373
Db      1281 CTGGTTCAGGCTTAGTCAC--CATTAAGCGCAGAGAGGCTTAAGGTATTAACCAAC 1337
OY      1374 CCTTAACCTATGCTGTGGGTATCAGACTGAGTGTGATGTGTCACAGCTGCTCAG 1433
Db      1338 TGAGTGCCTCTTCTCTCTCAGCCCAAGAGAGGGCTTGTGTATATCTATCTGA 1397
OY      1434 CAGAGAGCTTTTGTGTCATGCTGTCTAGAAACAGACTGGGAACTTATGTAG 1493
Db      1398 CCTTACACTGAGCCTTAAGAAAGAAAGATATCTTGTCTTGAATATCTTAGG 1457
OY      1494 CAGCAGATCCCAAGCTGAAGAGGTA--TTGCTCTCTCTTCTTCTTGTATCTTCG 1550
Db      1458 AACCTTACAGAGCCTTGTGACCTGTCAAGCAAGACTTCTGTATACCTTGGCTCTTA 1517
OY      1551 TCTGGGACACTTCAAGAGCTTGTGGCCTGGAGGCTTATTAAGCAGCAGCATATCAG 1610
Db      1518 CCTGTGCATACCTTGGAGACTAGGCTGTGGAGGCATACGAGACAGCAACAGCATCG 1577
OY      1611 TGGATTTATCCATTAACCTCCCTGTCCATCTTGGCCCAATGGGGAATGATCTTTCAC 1670
Db      1578 TGGAGTGTCTGTAACTCCCTGTCCACATCTTCCGAAAGAAAGGGAATTAATA 1637
OY      1671 CAAAGAGCTCAGCAGATTTTCCACAGATGGA-----ATTGTGAGCCCT 1717
Db      1638 GCCAAGGAAAGCGGCTTTCACCAAGAGCTCCCAAACTTCCACAGAGCATTAACCC 1697
OY      1718 TGGAGTTCCCAATGGGATTCAGAGAGAGAGTGGGAACAAGTTGGATCTTATATGA 1777
Db      1698 TCAGAGTTCCAGAGGATTTGAGAGAGAGAGCGGGGTGGTGTATCT-CTGAATC 1756
OY      1778 GCTTACCATACAGCTATGCTATCAAGAAATATGAACAATAATCTCTGACAAAGAGA 1837
Db      1757 TCTGCCATTTGACATATCAGTAATCAGAAACAGAGCAAAA--CAACACAAAGAGGC 1813
OY      1838 AGCTCTTAGTTCACAAGGTGCTGGCTTGTAT 1870
Db      1814 AAGACTCAAGTTCAGAGCGGTGTGTGTATGAT 1846

```

```

RESULT 3
AL709927      668 bp      mRNA      1linear      EST 12-JUN-2003
LOCUS      DKFZp68602364_r1.686 (synonym: hlcc3) Homo sapiens cDNA clone
DEFINITION      DKFZp68602364 5', mRNA sequence.
ACCESSION      AL709927
VERSION      AL709927.1 GI:19693282
KEYWORDS      EST.
SOURCE      Homo sapiens (human)
ORGANISM      Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
REFERENCE      1 (bases 1 to 668)
AUTHORS      Bahr, A., Lauber, J., Mewes, H.W., Well, B., Amid, C., Osanger, A., Fodo

```

```

FEATURES
source
    1..668
        /organism="Homo sapiens"
        /mol_type="mRNA"
        /db_xref="taxon:9606"
        /clone="DKFZp68602364"
        /tissue.type="human skeletal muscle"
        /dev_stage="adult"
        /lab_host="DH10B"
        /clone_lib="686 (synonym: hlcc3)"
        /note="Vector: pTriplex2; Site 1: SfiI; Site 2: SfiIb;
        cDNA-collection"
BASE COUNT      132 a      178 g      154 t      2 others
ORIGIN
Query Match      31.7%; Score 647.6; DB 9; Length 668;
Best Local Similarity 99.1%; Pred. No. 2,6e-160;
Matches 661; Conservative 0; Mismatches 5; Indels 1; Gaps 1;
OY      36 GGGTAGAGAGAAAGCAGATGGCGCGGTAGAGTGAAGCTCTCAAAAGCAGCAGGA 95
Db      2 GGGAGAGAGAGAAAGCAGATGGCGCGGTAGAGTGAAGCTCTCAAAAGCAGCAGGA 61
OY      96 GCCAAGCAGCAAGGCTTCCACTTACAGCAATGCTACTGCTTAAAAAATGAAGCTCC 155
Db      62 GCCAAGCAGCAAGGCTTCCACTTACAGCAATGCTACTGCTTAAAAAATGAAGCTCC 121
OY      156 TGTCTTGTGTTTCCAGATGGCCATCTGTGCTTATTTCTTCCATGTATACGCACA 215
Db      122 TGTCTTGTGTTTCCAGATGGCCATCTGTGCTTATTTCTTCCATGTATACGCACA 181
OY      216 ACATCAGCTCCCTGTATGAAGGACACCGGAGCGATGCAAGCTGTCTCTCTT 275
Db      182 ACATCAGCTCCCTGTATGAAGGACACCGGAGCGATGCAAGCTGTCTCTCTT 241
OY      276 CCTGGGCTGTGGCTCTTCTTGTGGGAGCTTTTGGGAGCACCAGATGTTTCT 335
Db      242 CCTGGGCTGTGGCTCTTCTTGTGGGAGCTTTTGGGAGCACCAGATGTTTCT 301
OY      336 ACCTGATGAGAGCGCGCTGGCAGTGTGATGATCTTCAAGAGAGACACCGCTGATGC 395
Db      302 ACCTGATGAGAGCGCGCTGGCAGTGTGATGATCTTCAAGAGAGACACCGCTGATGC 361
OY      396 TGCACATGGCTGTGGGATCTGATACGGGCGCTTCTTGTGCGCATGAGCGTCTTGG 455
Db      362 TGCACATGGCTGTGGGATCTGATACGGGCGCTTCTTGTGCGCATGAGCGTCTTGG 421
OY      456 ATGCCATCATGGAACCTGTCTCCCGGAGAGCAGTCCAGCCTCTTCAAGTGGGAGAACGCC 515
Db      422 ATGCCATCATGGAACCTGTCTCCCGGAGAGAGTCCAGCCTCTTCAAGTGGGAGAACGCC 481
OY      516 GGGCCCTGTGTCTGACCTGCTGTGACATATCCCAAAAGTGAATCATCCCGG-6 574
Db      482 GGGCCCTGTGTCTGACCTGCTGTGACATATCCCAAAAGTGAATCATCCCGG-6 541
OY      575 GGTCACTGAGGCTCTGTGACAGTCAACAGCCCTTGTGAGTGGTGAAGAGCGTGGCGC 634
Db      542 GGTCACTGAGGCTCTGTGACAGTCAACAGCCCTTGTGAGTGGTGAAGAGCGTGGCGC 601

```

QY 635 TCCTACAGCCAGTGTGCTCAAGAGAGTGGCTTTCTTCAACCTGCAGTCCCTTACCCG 694
 DB 602 TCTACAGCCAGTGTGCTCAAGAGAGTGGCTTTCTTCAACCTGCAGTCCCTTACCCG 661
 QY 695 CTGCTGA 701
 DB 662 CTGCTGA 668

RESULT 4
 ID HSM082755 standard; RNA; EST; 583 BP.
 AC BX490456;
 SV BX490456.1
 XX 09-MAY-2003 (Rel. 75, Created)
 DT 09-MAY-2003 (Rel. 75, Last updated, Version 1)
 XX Homo sapiens mRNA; EST DKFZp686G0680_r1 (from clone DKFZp686G0680)
 DE EST; expressed sequence tag.
 XX Homo sapiens (human)
 OS Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia;
 OC Eutheria; Primates; Catarrhini; Homnidae; Homo.
 XX [1]
 RP 1-583
 RA Ansoyge W., Krieger S., Regiert T., Rittmuller C., Schwager B.,
 RA Mewes H.W., Well B., Amlid C., Osanger A., Fobo G., Han M., Wiemann S.;
 RT Submitted (07-MAY-2003) to the EMBL/GenBank/DBJ databases.
 RL MIPS, Ingolstaedter Landstr.1, D-85764 Neuberg, GERMANY
 XX This is the 5' sequence of the clone insert
 CC Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 CC Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
 CC sequenced by EMBL (European Molecular Biology Laboratories,
 CC Heidelberg/Germany) within the cDNA sequencing consortium of
 CC the German genome project.
 CC No s1 sequence available.
 CC This clone (DKFZp686G0680) is available at the RZPD in Berlin.
 CC Please contact the RZPD: Ressourcezentrum, Heubnerweg 6,
 CC 14059 Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de

Key Location/Qualifiers
 FT 1..583
 FT /db_xref="taxon:9606"
 FT /mol_type="mRNA"
 FT /organism="Homo sapiens"
 FT /clone="DKFZp686G0680"
 FT /clone_1lb="686 (synonym: hlcc3). Vector psport1_sfi; host
 FT DH10B; sites SfiI + SfiIb"
 FT /dev_stage="adult"
 FT /tissue_type="CDNA-collection"
 XX Sequence 583 BP; 120 A; 172 C; 154 G; 137 T; 0 other;

Query Match 28.5%; Score 581.4; DB 2; Length 583;
 Best Local Similarity 99.8%; Pred. No. 9.2e-143;
 Matches 583; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 41 GAGGAGAAAGCCATGCGCCGCTAGCACTGAGCTCTCAAAAAGCAGAGGAAGCCCA 100
 DB 1 GAGGAGAAAGCCATGCGCCGCTAGCACTGAGCTCTCAAAAAGCAGAGGAAGCCCA 60
 QY 101 AGCCACAAGCTCTTCCCTCAGACACATGCTACTGCTTAAAAAATGAAGCTCTCTG 160
 DB 61 AGCCACAAGCTCTTCCCTCAGACACATGCTACTGCTTAAAAAATGAAGCTCTCTG 120

QY 161 TTTCGTTTCCAGATGGCATCTTGGCTTATCTTCCAGATGTACAGCACACATC 220
 DB 121 TTTCGTTTCCAGATGGCATCTTGGCTTATCTTCCAGATGTACAGCACACATC 180
 QY 221 AGCTCCCTGTATGAAGGACACAGCCGAGGAGCATGACGTGTCTTCTCTCTG 280
 DB 181 AGCTCCCTGTATGAAGGACACAGCCGAGGAGCATGACGTGTCTTCTCTCTG 240
 QY 281 CGCTGTGGCTCTTCTTGTGGGAGCCTTTTGGGAGCACCAGATGTTTCTACCTG 340
 DB 241 CGCTGTGGCTCTTCTTGTGGGAGCCTTTTGGGAGCACCAGATGTTTCTACCTG 300
 QY 341 ATGAGGCGCCGCTGGACAGTGTGATACCTTCAAGAGAGACCGCTGATCTCTAC 400
 DB 301 ATGAGGCGCCGCTGGACAGTGTGATACCTTCAAGAGAGACCGCTGATCTCTAC 360
 QY 401 ATGCTGTGGGAGATGATACGAGGCGCTCTTGTGCGACATGACGCTTTGATGCC 460
 DB 361 ATGCTGTGGGAGATGATACGAGGCGCTCTTGTGCGACATGACGCTTTGATGCC 420
 QY 461 TACATGACCTGTGCTCCCGAGACAGTCCAGCTCTTTCAGTGGAGAACAGCCGGCC 520
 DB 421 TACATGACCTGTGCTCCCGAGACAGTCCAGCTCTTTCAGTGGAGAACAGCCGGCC 480
 QY 521 CTGTGTTCTGACCTGCTGTGACATCATCCACAGATGAAATCATCCCGGGCTCAC 580
 DB 481 CTGTGTTCTGACCTGCTGTGACATCATCCACAGATGAAATCATCCCGGGCTCAC 540

QY 581 TGCAGGCTCTGTGACATCAACAGCCCTTGTAGGTGTGGAGA 623
 DB 541 TGCAGGCTCTGTGACATCAACAGCCCTTGTAGGTGTGGAGA 583

RESULT 5
 BU684395/c 669 bp mRNA linear EST 07-Oct-2002
 LOCUS
 DEFINITION UI-CF-ENO-aco-f-08-0-UI.s1 UI-CF-ENO Homo sapiens cDNA clone
 ACCESSION BU684395
 VERSION BU684395.1 GI:23537302
 KEYWORDS EST.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 669)
 Bernaldo,M.F., Lennon,G. and Soares,M.B.
 Normalization and subtraction: two approaches to facilitate gene
 discovery
 Genome Res. 6 (9), 791-806 (1996)
 JOURNAL MEDLINE
 PUBMED 8889548
 COMMENT Contact: McCray, PB
 McCray Lab
 University of Iowa
 2024 University of Iowa Med Labs, Iowa City, IA 52242, USA
 Tel: 319 356 4866
 Fax: 319 356 7171
 Email: paul-mccray@uiowa.edu
 Tissue Procurement: Dr. M. J. Welsh, University of Iowa
 CDNA Library Preparation: Dr. M. Bento Soares, University of Iowa
 DNA Sequencing by: Dr. M. Bento Soares, University of Iowa
 Clone Distribution: Researchers may obtain clones from Research
 Genetics (www.resgen.com) or from Open Biosystems
 Seq primer: M13 FORWARD
 POLYA=yes

FEATURES
 source location/Qualifiers
 1..669
 /organism="Homo sapiens"
 /mol_type="mRNA"

ORIGIN

Query Match 24.8% Score 507: DB 12: Length 593:
 Best Local Similarity 97.4%: Pred. No. 4.9e-123;
 Matches 568: Conservative 0: Mismatches 10: Indels 5: Gaps 5:

QY 1382 ACATGCTGTGGGATATCAGACATGAGTGTGATGTCACACAGCTGCTCAAGAGAGA 1441
 DB 593 ACATGCTGTGGGATATCAGACATGAGTGTGATGTCACACAGCTGCTCAAGAGAGA 534
 QY 1442 CTTTGTGTCCATGCTGTGTCTAGAAACAGACTGGGGAACCTTATGTAGCAGCAAT 1501
 DB 533 CTTTGTGTCCATGCTGTGTCTAGAAACAGACTGGGGAACCTTATGTAGCAGCAAT 474
 QY 1502 CCCACGAGTGAACAGGATATGCTCTTCTTCTTGTATGTCCTGCTGCTGGGAGAC 1561
 DB 473 CCCACGAGTGAACAGGATATGCTCTTCTTCTTGTATGTCCTGCTGCTGGGAGAC 414
 QY 1562 TTCAGACATTTGTGGCTGTGGAGCCCTATTAAAGCAGACAGATATCAGTGAATGATC 1621
 DB 413 TTCAGACATTTGTGGCTGTGGAGCCCTATTAAAGCAGACAGATATCAGTGAATGATC 354
 QY 1622 CATTAACCTCCCTGTCCACATCTTGCCCAATGGGGAATGATCTTCAACCAAGAGCTCA 1661
 DB 353 CATTAACCTCCCTGTCCACATCTTGCCCAATGGGGAATGATCTTCAACCAAGAGCTCA 294
 QY 1682 CCAGCATTTTCCACAGAGATGCGAATCTGTAGCCCTTGGAGTTGCCAATGGGATTCAGG 1741
 DB 293 CCAGCATTTTCCACAGAGATGCGAATCTGTAGCCCTTGGAGTTGCCAATGGGATTCAGG 235
 QY 1742 AAGGAAGTGGGAAACAGGTTGGATGCTTACTTATGAGCTTGACCAT-ACAGCTATCGGTA 1800
 DB 234 AAGGAAGTGGGAAACAGGTTGGATGCTTACTTATGAGCTTGACCATGACAGCTATCGGTA 175
 QY 1801 ATCGAATATNGAACAACAAATCTCTG-ACAAAAGCAGAGCTTAAAGTTACAGAGTGC 1859
 DB 174 ATCGAATATNGAACAACAAATCTCTG-ACAAAAGCAGAGCTTAAAGTTACAGAGTGC 115
 QY 1860 CTGGGCTTGAATTAATATCA-TTCCCTTGTGATTTTCCATATACATAGAAATCTTGA 1918
 DB 114 CTGGGCTTGAATTAATATCA-TTCCCTTGTGATTTTCCATATACATAGAAATCTTGA 55
 QY 1919 CCTGTGAAGCTGGCATCTGTTAACTAAATCCCAATTA 1961
 DB 54 CCTGTGAAGCTGGCATCTGTTAACTAAATCCCAATTA 13

RESULT 7

BM129080

LOCUS

1417c04.y1 Melton Normalized Human Islet 4 N4-HIS 1 Homo sapiens
 N-ACETYLGLUCOSAMINE 6-O-SULFOTRANSFERASE. ; mRNA sequence.

DEFINITION

BM129080 571 bp mRNA linear EST 12-MAR-2002
 BM129080.1 GI:17123632

ACCESSION

BM129080.1 GI:17123632

VERSION

EST.

KEYWORDS

Homo sapiens (human)

SOURCE

Homo sapiens

ORGANISM

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE

1 (bases 1 to 571)

AUTHORS

Melton,D., Brown,T.J., Kenty,G., Permutt,A., Lee,C., Kaestner,K.,
 Lemishka,I., Searles,M., Brestelli,J., Gradwohl,G., Clifton,S.,
 Hillier,L., Marra,M., Page,D., Wylie,T., Martin,J., Blisstein,A.,
 Schmitt,A., Theising,B., Ritzer,E., Ronko,I., Bennett,J., Cardenas,
 M., Gibbons,M., McCann,R., Cole,R., Tsagarelis,V., Williams,T.,
 Jackson,Y., and Bowers,Y.

TITLE

Endocrine Pancreas Consortium

JOURNAL

Unpublished

COMMENT

Other ESTs: 1417c04.x1
 Contact: Douglas Melton, Klaus H. Kaestner, & Hiroshi Inoue
 Endocrine Pancreas Consortium
 Harvard University, Howard Hughes Medical Institute

FEATURES

source

Dept of Molecular and Cellular Biology, 7 Divinity Ave, Cambridge,
 MA 02138
 Tel: 617-495-1812
 Fax: 617-495-8557
 Email: dmelton@molbio.harvard.edu
 Library was constructed by Dr. Douglas Melton DNA sequencing by:
 Washington University Genome Sequencing Center For information on
 obtaining a clone please contact: Juliana Brown
 (brownj@fas.harvard.edu) This sequence now available from the IMAGE
 consortium, for clone orders contact: info@image.llnl.gov
 High quality sequence stop: 434.

Location/Qualifiers

1..571

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:567655"

/sex="Both"

/tissue_type="Islets of Langerhans"

/dev_stage="Adult"

/lab_host="DH10B"

/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"

/note="Organ: Pancreas; Vector: pSPORT1; Site: 1; Not 1;
 Site 2: Sal 1; Starting library constructed using
 SuperScript Plasmid Library kit (Life Technologies). cDNA
 made by oligo-dT priming. Size-selected by column
 fractionation; average insert size 1.08 kb. Library was
 amplified once on solid support and plasmid DNA from
 library was prepared. The library DNA was normalized by
 method #4 from Bonaldo, Lennon, and Soares 1996 Genome
 Research 6:791-806; 0.5 microgram single-stranded library
 plasmid DNA was mixed with 5 micrograms PCR product
 representing library inserts and hybridized to an EcoT of
 20. Single-stranded (unhybridized) plasmids were isolated
 by hydroxyapatite chromatography and used to make this
 library."

BASE COUNT

109 a 169 c 143 g 150 t

ORIGIN

Query Match

24.5% Score 499.8: DB 12: Length 571;

Best Local Similarity 96.6%: Pred. No. 4e-121;

Matches 504: Conservative 0: Mismatches 7: Indels 0: Gaps 0;

QY 94 AAGCCAAAGCCAGAGGCTCTCCACTTCAGACATGCTACTGCTTAAAAATGAAGCT 153
 DB 61 AAGCCCGCTTGCAAGGCTTCCACTTCAGACATGCTACTGCTTAAAAATGAAGCT 120
 QY 154 CCGCTGTTCTGTGTTTCCAGATGCGCATCTTGCTATCTTCCATGTAAGCA 213
 DB 121 CCGCTGTTCTGTGTTTCCAGATGCGCATCTTGCTATCTTCCATGTAAGCA 180
 QY 214 CAATATAGCTCCCTGTCTATGAGAGCAGACACCCGAGGAGCATGACGCTGCTGCTC 273
 DB 181 CAATATAGCTCCCTGTCTATGAGAGCAGACACCCGAGGAGCATGACGCTGCTGCTC 240
 QY 274 TTCTGGGAGCTGCTGCTCTTCTTGTGGGAGCAGTGGAGACCCAGATGTTT 333
 DB 241 TTCTGGGAGCTGCTGCTCTTCTTGTGGGAGCAGTGGAGACCCAGATGTTT 300
 QY 334 CTACCTGATGAGGCGCGCTGCGACGCTGTGATGATCTTCAAGAGAGACCGCTGAT 393
 DB 301 CTACCTGATGAGGCGCGCTGCGACGCTGTGATGATCTTCAAGAGAGACCGCTGAT 360
 QY 394 GCTGACATGCTGCTGCGGATGATGATGAGGCGCGCTCTTCTTGTGGCAGATGACGCTT 453
 DB 361 GCTGACATGCTGCTGCGGATGATGATGAGGCGCGCTCTTCTTGTGGCAGATGACGCTT 420
 QY 454 TGATGCTTACATGAGACCTGCTCCGAGAGACATGACGCTCTTCAAGTGGAGAACAG 513
 DB 421 TGATGCTTACATGAGACCTGCTCCGAGAGACATGACGCTCTTCAAGTGGAGAACAG 480
 QY 514 CCGGCGCTGTGTTCTGACCTGCTGTGACATGATCCACAGATGAATCATCCCGC 573

Dd		481	CCGGGCGCCTGTTCGTGCACCTGCCTGTGCATCATCATCCACAAGATGAATATCATCCCCG	540
Oy		574	GGCTCACTGCAGAGCTCCTGTGCATCAACAG	604
Dd		541	GGCTCACTGCAGAGCTCCTGTGCATCAACAG	571
RESULT 8				
Bf878439				
LOCUS				
DEFINITION	Bf878439..	680 bp	mRNA	linear EST 17-JAN-2001
ACCESSION	MRO-ET0109-191100-002-h06 ET0109	Homo sapiens cDNA,	mRNA sequence.	
VERSION	Bf878439.1	GI:12268569		
KEYWORDS	EST.			
SOURCE	Homo sapiens			
ORGANISM	Homo sapiens (human)			
EREFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
AUTHORS	Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
	1 (bases 1 to 680)			
	Dias Neto,E., Garcia Correia,R., Verjovski-Almeida,S., Britones,M.R.,			
	Nagai,M.A., da Silva,W. Jr., Zaog,M.A., Bordin,S., Costa,F.F.,			
	Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S. Simpson,D.H.,			
	Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare			
	,M.J., Soares,F., Brentani,R.K., Reis,L.F., de Souza,S.U. and			
	Simpson,A.J.			
TITLE	Shotgun sequencing of the human transcriptome with ORF-expressed			
JOURNAL	sequence tags			
MEDLINE	Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)			
PUBMED	20202653			
COMMENT	10737800			
	Contact: Simpson A.J.G.			
	Laboratory of Cancer Genetics			
	Ludwig Institute for Cancer Research			
	Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,			
	Brazil			
	Tel: +55-11-2704922			
	Fax: +55-11-2707001			
	Email: asimpson@ludwig.org.br			
	This sequence was derived from the FAPESP/LICR Human Cancer Genome			
	Project. This entry can be seen in the following URL			
	(http://www.ludwig.org.br/scripts/gethtml2.pl?fl=MR06t2-MRO-ET0109-			
	191100-002-h06&t3=2000-11-19&t4=1)			
	Seq primer: puc 18 forward			
	High quality sequence start: 17			
	High quality sequence stop: 641.			
FEATURES				
Source	Location/Qualifiers			
	1..680			
	/organism="Homo sapiens"			
	/mol_type="mRNA"			
	/db_xref="taxon:9606"			
	/dev_stage="Adult"			
	/clone_lib="ET0109"			
	/note="Organ: Lung_tumor; Vector: puc18; Site_1: SmaI;			
	Site_2: SmaI; A mini-library was made by cloning products			
	derived from ORBESTES PCR (U.S. Letters Patent application			
	No. 196,716 - Ludwig Institute for Cancer Research)			
	profiles into the pUC 18 vector. Reverse transcription o			
	tissue mRNA and cDNA amplification were performed under			
	low stringency conditions."			
BASE COUNT	168 a 164 c 163 g 164 t 1 others			
ORIGIN				
Query Match	24.3%; Score 496.2; DB 10; Length 680;			
Best Local Similarity	93.5%; Pred. No.3,7e-120;			
Matches	572; Conservative 0; Mismatches 34; Indels 6; Gaps 5;			
Oy	1328 CAGTCACTTTCTGTGATGCTTGAGCGCTTGCT-ACATCTGAGGCTTAACATGATG	1386		
Dd	17 CAGTCACTTTCTGTGATGCTTGAGCGCTGCCCTAATCATCTGTGAGCTTAACATG	76		
Oy	1387 TCTGTGGGTACACACT--GAGTGTGATTGTGTGCACAGCGTCACAGCAAGAAGACTT	1444		
Dd	77 TCTGTGGGTACACACTGAGTGTGGAGTGTGTCCACAGCGTCACAGCAAGAAGACTT	136		

[illegible]

FEATURES
sourceLocation/Qualifiers
1. .553

/organism="Homo sapiens"

/mol_type="mRNA"

/db_xref="taxon:9606"

/clone="IMAGE:567635"

/sex="Both"

/tissue_type="Islets of Langerhans"

/dev_stage="Adult"

/lab_host="DH10B"

/clone_lib="Melton Normalized Human Islet 4 N4-HIS 1"

/note="Organ: Pancreas; Vector: pSPORT1; Site_1: Not 1; Site_2: Sal 1; Starting library constructed using SuperScript Plasmid Library kit (Life Technologies). cDNA made by oligo-dT priming. Size-selected by column fractionation: average insert size 1.08 kb. Library was amplified once on solid support and plasmid DNA from library was prepared. The library DNA was normalized by method #4 from Bonaldo, Lennon, and Soares 1996 Genome Research 6:791-806; 0.5 microgram single-stranded library plasmid DNA was mixed with 5 micrograms PCR product representing library inserts and hybridized to an EcoT of 20. Single-stranded (unhybridized) plasmids were isolated by hydroxyapatite chromatography and used to make this library."

BASE COUNT

163 a 111 c 127 g 152 t

ORIGIN

Query Match 23.3%; Score 476; DB 12; Length 553;
Best Local Similarity 97.3%; Pred. No. 8.1e-115;

Matches 537; Conservative 0; Mismatches 10; Indels 5; Gaps 5;

QY 1486 TATGTGAGCAGACATCCACAGTGAAGAGGATATGCTCTCTCTTTCTTGATCT 1545

Db 553 TATGTGAGCAGACATCCACAGTGAAGAGGATATGCTCTCTCTTTCTTGATCT 494

QY 1546 TCCTGTCTGGCAGACTTCAGAGACTTTGGCTGAGGCGCTATTAGACGACACAGT 1605

Db 493 TCCTGTCTGGCAGACTTCAGAGACTTTGGCTGAGGCGCTATTAGACGACACAGT 434

QY 1606 ATCAGTGGAAATGATCCATAAAGCTCCCTGTCACATCTTCCCAATGGGGAATGATCT 1665

Db 433 ATCAGTGGAAATGATCCATAAAGCTCCCTGTCACATCTTCCCAATGGGGAATGATCT 374

QY 1666 TTCACAAAGAGCTCACAGATTTTCCACAGAGATGCGAATTCAGAGCCCTGGAGTTC 1725

Db 373 TTCACAAAGAGCTCACAGATTTTCCACAGAGATGCGAATTCAGAGCCCTGGAGTTC 314

QY 1726 CCAATGGGATTCAGAGAGAGTGGCAAGAGTGGATGCTTACTTATGAGCTTGACC 1785

Db 313 CCAATGGGATTCAGAGAGAGTGGCAAGAGTGGATGCTTACTTATGAGCTTGACC 255

QY 1786 AT-ACAGCTATCGTATCAGAAATATGAAAACAAATCTCTG-ACAAAAGACAGCTCT 1843

Db 254 ATCAGCTATCGTATCAGAAATATGAAAACAAATCTCTG-ACAAAAGACAGCTCT 195

QY 1844 TAACTGACAAAGTGGCTGGCTGATTTGAATATCA-TTTCCTTTGCAATTTCCATTT 1902

Db 194 TAACTGACAAAGTGGCTGGCTGATTTGAATATCA-TTTCCTTTGCAATTTCCATTT 135

QY 1903 ACATAGAAACTTGTACCTGTGAAGTCCATCTGTATATCTAAATTTCCCAATGAG 1962

Db 134 ACATAGAAACTTGTACCTGTGAAGTCCATCTGTATATCTAAATTTCCCAATGAG 76

QY 1963 GTTCTGTTAGAAATGCTCTTTTATGCTTCTTAATATTTAGAGTAATGCTTATTTT 2022

Db 75 GTTCTGTTAGAAATGCTCTTTTATGCTTCTTAATATTTAGAGTAATGCTTATTTT 16

QY 2023 ATGGGATCTTAA 2034

Db 15 ATGGGATCTTAA 4

RESULT 10
BM128831/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

COMMENT

JOURNAL

Db 493 TCCTGCTGGGACAGCTTACAGACCTTTGCTGGCGGAGCCCTATTAAAGCAGACAGCT 434
 QY 1606 ATCAGTGAATGATCCATTAACCTCCCTGTCACATCTTGCCCAATGGGAATGATCT 1665
 Db 433 ATCAGTGAATGATCCATTAACCTCCCTGTCACATCTTGCCCAATGGGAATGATCT 374
 QY 1666 TTCACCAAGAGCTCACCGCATTTTCCACAGAGATGCCAATTTGAGCCCTTGAAGTTC 1725
 Db 373 TTCACCAAGAGCTCACCGCATTTTCCACAGAGATGCCAATTTGAGCCCTTGAAGTTC 314
 QY 1726 CCAATGATTCAGAGAGAGAGTGGGAACAGGTTGATGCTTACTTATGAGCTTGACC 1785
 Db 313 CCAATGATTCAGAGAGAGAGTGGGAACAGGTTGATGCTTACTTATGAGCTTGACC 255
 QY 1786 AT-ACAGCTATCGGTATATGAGAAATATGAAACAAATCTCTG-ACAAAGAGCAAGCTCT 1843
 Db 254 ATCAGCTATCGGTATATGAGAAATATGAAACAAATCTCTGACAAAGAGCAAGCTCT 195
 QY 1844 TAAGTTCACAGAGTGGCTGGGCTTGATTTGAATATCA-TTTCCTTTGCAATTTTCCATTT 1902
 Db 194 TAAGTTCACAGAGTGGCTGGGCTTGATTTGAATATCA-TTTCCTTTGCAATTTTCCATTT 135
 QY 1903 ACATGAAAACCTTTGACCTGTGTAACCTGCAATCTGTAATTAATAATTCACAAATTAAG 1962
 Db 134 ACATGAAAACCTTTGACCTGTGTAACCTGCAATCTGTAATTAATAATTCACAAATTAAG 76
 QY 1963 GTTCTGTTTGAAGATGCTCTTTTATGCTTCTTAATTAATTAAGCAATTAATTTTCAATTTT 2022
 Db 75 GTTCTGTTTGAAGATGCTCTTTTATGCTTCTTAATTAATTAAGCAATTAATTTTCAATTTT 16
 QY 2023 ATGGGATCCTAA 2034
 Db 15 ATGGGATCCTAA 4
 RESULT 11
 BG963298 783 bp mRNA linear EST 12-JUN-2001
 LOCUS 602827716F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4982643 5',
 DEFINITION mRNA sequence.
 ACCESSION BG963298
 VERSION BG963298.1 GI:14350935
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)
 Unpublished
 CONTACT: Robert Strausberg, Ph.D.
 Email: cgabbs-r@mail.nih.gov
 Tissue Procurement: Jeffrey E. Green, M.D.
 cDNA Library Preparation: Life Technologies, Inc.
 cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
 DNA Sequencing by: Incyte Genomics, Inc.
 Clone Distribution: MGC clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
 http://image.llnl.gov
 Plate: L1AM10985 row: P column: 04
 High quality sequence stop: 741.
 Location/Qualifiers
 1..783
 /organism="Mus musculus"
 /mol_type="mRNA"
 /strain="FVB/N"
 /db_xref="taxon:10090"
 /clone="IMAGE:4982643"
 /lab_host="DH10B (T1 phage-resistant)"
 /clone_id="NCI_CGAP_Co24"
 /note="Organ: colon; Vector: pCMV-SPORT6; Site: J; NotI;
 Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.

FEATURES
 source

BASE COUNT 141 a 236 c 226 g 180 t
 ORIGIN

Query Match 18.9%; Score 385.2; DB 12; Length 783;
 Best Local Similarity 77.2%; Pred. No. 1e-90;
 Matches 545; Conservative 0; Mismatches 153; Indels 8; Gaps 6;

QY 128 ATGCTAGTGCCTTAAAGAAAGCTGCTGCTTTCTGTTTCCAGATGGCATCTTG 187
 Db 67 ATGCTAGTGCCTTAAAGAAAGCTGCTGCTTTCTGTTTCCAGATGGCATCTTG 126
 QY 188 GCTATTTCTTCCATGATGACACCAACATCAATGCTGCTGCTTATGAGGACAGCC 247
 Db 127 GCTCTTTCATCATATGTCCTGTCACAGACACTTTCACAGAGGAGGATCCAGAG- 185
 QY 248 GAGCGATGACAGTGGTGTCTGCTTCTTCTGCGGCTGCTGCTTCTTTTGGGGGAG 307
 Db 186 --GCCGTGATGTGCTGCTGCTGCTTCTTCTGCGGCTGCTGCTTCTTTTGGGGGAG 243
 QY 308 CTTTGGGACAGACCAAGATGTTTCTACCTGATGAGCCGCGCTGGCAGCTGGATG 367
 Db 244 CTTTGGGACAGACCAAGATGTTTCTACCTGATGAGCCGCGCTGGCAGCTGGATG 303
 QY 368 ACCTTCAAGCAGACACCGCTGATGCTGACATGCTGTGGGGATGTATAGGGGCC 427
 Db 304 ACTTTCACAGCAGACACCGCTGATGCTGACATGCTGTGGGGATGTATAGGGGCC 363
 QY 428 GTCTCTTGTGGACATGAGCGCTTGTATGCTTATGAAACCTGCTGCGGAGACAG 487
 Db 364 GTCTCTTGTGGACATGAGCGCTTGTATGCTTATGAAACCTGCTGCGGAGACAG 423
 QY 488 TCCAGCCTTCTTCAAGTGGAGAAACACCGGCGCTGTCTTCTGACCTGCTGTGATC 547
 Db 424 TCCAGCCTTCTTCAAGTGGAGAAACACCGGCGCTGTCTTCTGACCTGCTGTGATC 483
 QY 548 ATCCCAAGATGAATATATCCCGGGGCTCACTGACAGGCTCTGTGACGTCAACAGCC 607
 Db 484 TTCCCTGCGCCAGAGTACAGCTCACCAAGCACTGCAACCTGCTGCGGTGACAGCGCC 543
 QY 608 TTTGAGTGTGGAGAAAGCTGCGCTCTCTACAGCAGCAGTGGTGTCA-GGAGGTGG 666
 Db 544 TTGATATGATGAGAGAGGCTCTGCTCTACAGGCTTGTGTACTCAAGAGGTGG 603
 QY 667 -CTTCTCAACCTGACAGTCCCTTACCCGCTGCTGTAAGAACCCCTCTCAACCTG-CAT 724
 Db 604 TATCTCTAGCGTCGAGGCGCTCTATTCACAGGCTCTCTCAACCTGCTCAACCTGCCAC 663
 QY 725 ATGCTGACCTGCTGCGGAGACCCCGGCGCTTCTGCTTCCGAGAGACGCAAGAGGA 784
 Db 664 GTCTGTCACCTGCTGCGGAGACCCCGGCGCTTCTGCTTCCGAGAGACGCAAGAGGA 722
 QY 785 GATCTCATATGTA-CAGTGCATTTGATGAGGAGCAGATGAGCAA 829
 Db 723 GAACCTGCTGTTACAGTCAATCTGCTGTAAGGCGACATTTGAAA 788
 RESULT 12
 BG964671 695 bp mRNA linear EST 12-JUN-2001
 LOCUS 602831875F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4986258 5',
 DEFINITION mRNA sequence.
 ACCESSION BG964671
 VERSION BG964671.1 GI:14352308
 KEYWORDS EST.
 SOURCE Mus musculus (house mouse)
 ORGANISM Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 NIH-MGC http://mgi.nci.nih.gov/
 National Institutes of Health, Mammalian Gene Collection (MGC)

JOURNAL
COMMENT

Unpublished
Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM10995 row: f column: 19
High quality sequence stop: 684.
Location/Qualifiers

FEATURES
source

1. .695
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4986258"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 118 a 216 c 195 g 166 t
ORIGIN

Query Match 18.0%; Score 368.6; DB 12; Length 695;
Best Local Similarity 74.9%; Pred. No. 2.6e-86;
Matches 501; Conservative 0; Mismatches 164; Indels 4; Gaps 3;

111 TCTTCCACTGACATGCTACTGCTAAATAAAGCTGCTGTTT 170
19 TCTCTTCTTCCGAGATGATGCTGTGAAGAAAGAGCTGCTGTTCTGGGTT 78
171 CCCAGATGCGCATCTTGGCTCTATCTTCCATGATGACGACACATCAGTCCCT 230
79 CCCAGATGCGCATCTTGGCTCTATCTTCCATGATGACGACACATCAGTCCCT 138
231 CTATGAAGCGACACCCGAGCGATGACGCTGCTGTCTTCTTCTGCGCTGCT 290
139 --AGGAGAGATCCAGAGCGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 196
291 CTCTTTTGGGGGAGCTTTTGGGAGACCCAGATGTTTCTACTGATGAGCCCG 350
197 CCTC-TTGTGAGGACCTTTTGGGAGACCCGAGATGTTTCTACTGATGAGCCCG 255
351 CCTGCGACGTGTGATGACCTTCAAGCAGACACCGCTGATGCTGCATGCTGTC 410
256 CCTGCGACGTGTGATGACCTTCAAGCAGACACCGCTGATGCTGCATGCTGTC 315
411 GGGATGATACGGGCGCTCTTGTGCGACATGAGCGTTTGTGATGCTTACATG 470
316 GGGATGATACGGGCGCTCTTGTGCGACATGAGCGTTTGTGATGCTTACATG 375
471 CTGTCCCGGAGACAGTCCAGCTTTTGTGAGGAGACAGCGGCGCTGTGTTG 530
376 CAGGCGCCCGGAGACAGTCCAGCTTTTGTGAGGAGACAGCGGCGCTGTGTTG 435
531 CACGTGCTGTGATGATCCACAGATGAATCATCCCGGAGCTCACTGAGGCTCC 590
436 CGCTGTGTGATGATCCCGTCCCGGAGATGATGATCCCAAGACTGCAAGCTGC 495
591 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 650
496 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 555
651 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 710
556 TGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 615
711 -CCCTCAACCTGATATGTCACCTGATCCGAGACCCCGGCGCTGTGCTCCGGA 769

DB 616 CCCCTCAACCTGACATGCTGCTGACATGCTGCTGACAGACCCCGGCTGCTGATCCCGG 675
OY 770 GAAGGCACA 778
DB 676 GGAGCCACA 684

RESULT 13

LOCUS BG966340 852 bp mRNA linear EST 12-JUN-2001

DEFINITION 602832826P1 NCI_CGAP_CO24 Mus musculus cDNA clone IMAGE:4987327 5',
mRNA sequence.

ACCESSION BG966340.1 GI:14353977

VERSION BG966340.1

KEYWORDS EST.

SOURCE Mus musculus (house mouse)

ORGANISM Mus musculus

REFERENCE NIH-MGC <http://mgc.nci.nih.gov/>,
National Institutes of Health, Mammalian Gene Collection (MGC)

AUTHORS Unpublished

TITLE Unpublished

JOURNAL Contact: Robert Strausberg, Ph.D.
Email: cgaabs-remail.nih.gov

COMMENT Tissue Procurement: Jeffrey E. Green, M.D.
CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
<http://image.llnl.gov>
Plate: LLM10998 row: c column: 08
High quality sequence stop: 784.
Location/Qualifiers

FEATURES
source

1. .852
/organism="Mus musculus"
/mol_type="mRNA"
/strain="FVB/N"
/db_xref="taxon:10090"
/clone="IMAGE:4987327"
/lab_host="DH10B (TI phage-resistant)"
/note="Organ: colon; Vector: pCMV-Sport6; Site_1: NotI;
Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
Average insert size 1.6 kb. Constructed by Life
Technologies. Note: this is a NCI_CGAP library."
BASE COUNT 176 a 248 c 244 g 184 t
ORIGIN

Query Match 17.8%; Score 364.4; DB 12; Length 852;
Best Local Similarity 74.3%; Pred. No. 3.5e-85;
Matches 502; Conservative 0; Mismatches 166; Indels 8; Gaps 3;

116 CACTTCAACACATGCTACTGCTTAAATAAAGCTGCTGTTTCTGTTTCCAG 175
125 CAGTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 184
176 ATGGCATCTTGGCTCTATCTTCCATGATGATGATGATGATGATGATGATG 235
185 GTATATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241
236 AAGGACAGCCCGAGCGATGATGATGATGATGATGATGATGATGATGATGAT 295
242 GAGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 301
296 -TTGTGGGAGAGCTTTTGGGAGACACCCAGATGTTTCTACTGATGAGCCGCTG 354
302 ATATGTGGAGAGCTTTTGGGAGACACCCAGATGTTTCTACTGATGAGCCGCTG 361
355 GCAGGTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 414

[illegible]

Query Match Similarity							17.8%	Score 363.4	DB 10	Length 965;
Best Local Similarity							76.2%	Pred. No. 6.7e-85		
Matches 537; Conservative							0	Mismatches 161;	Indels	Gaps 7;
Oy	625	GGCCTGCCGCTACAGCCACGTGGTGCTCAAGAAGGTGGCGTTCTTCAACTGCACGTGCAGTC	684							
Db	2	GGCCTGCCGCTCTCACGGCTCTCGTGGTAATCAAGAGAGGTGGCTTTTCTACACCTGCACGGC	61							
Oy	685	CCTTACC GGCTGCTGAAAGAACCCCCTCCCTCAACCTGATATTCGTGCACCTGGTCCGGGA	744							
Db	62	CCTCATTCACATGCTGCACGACCCCTTCCCTCAACCTGCACGTGTGCACCTGGTCCGAGA	121							
Oy	745	CCCCCGGGCCCTGTGTCCGTTCCCGAAGACGACAAAGGAGGATATCATGATGTACAGTACG	804							
Db	122	CCCCCGGGCCGTGTTCGATCCCGGAGACACACCCATATAACTCGTGGTTAGACGTCA	181							
Oy	805	CATTGTGATGGGGCAGCATGAGCAAATACTCAAGAGAGGACCAACCTACTATGTGAT	864							
Db	182	TA-TGTGTAGGGAGAGCAT-TGGAACAACATCAAGAGAGAMACAGCCCTTATTATGGCAT	239							
Oy	865	GCAGGTGATTCGGCAAGCCAGCTGGAGATATCAAGAACATTCACAGTCCCTGGCCAAGC	924							
Db	240	GAAAGTCACTCTGCAAAAMCCAGGTGGACATATGTCAAGGCATTCCAAAACCTCTCCGAAAGC	299							
Oy	925	CCGTGAGAAAGCTACCTGCTTGTGCCTATAGAGGACCTGGCTGAGCCCCCTGTGGCCCA	984							
Db	300	TCTGACAGAGCGCTACCTGTTCTGTAGAGTATGAGAGACCTGGTTGGGGCACCCCTGGCCCA	359							
Oy	985	GAC TTCCTCCGATATATGAATTCGTGGCATTTGGAATTTCTGCCCCCATCTTCAGACCTGGGT	1044							
Db	360	GAGCACCCAGCATATATAAAT-TGTGGGTTTGATTTTGTGCCCACTCCAAAACATGGGT	418							
Oy	1045	GCATPAACATCACCCGAGGCAAGGGCATGGTAGACACGCTTCCACATAAATGGCAGGGA	1104							
Db	419	TCACAATTTACACCCGCGCAAGGGCATGGTGCAGCATGCTTCCATACTAACGCCAGGAA	478							
Oy	1105	TGCCCTTAATGTCTCCACAGGCTTGGCGCTGTCTTGGCCCTA-TGMAAAGTTTTCTCGAC	1163							
Db	479	CGCCCTCAAGCTCTCTCGAGGGGTGGCTGTGCTTACCTTAACGAAAAAGTTTTCCAGC	538							
Oy	1164	TTTCAGAAAGCTGTGGCGATGCCAT-GAATTTGCTGGGCTACCGCCACGTGATGTGAA	1222							
Db	539	TTTCAGAAAGCTGTGGCGATGCCATGATTTTGTCTGGGATATCTCCAGGTGAGATCTCAA	598							
Oy	1223	CAAACAAGCAAAACCTGTGCTGGATCTTGCTTACCTGACGTGCTCCAGCAAAATC	1282							
Db	599	CAGAGAAC-AAGGCACATCTCCCTGATTTCTGTCTCTCTCCCATATCTTGGGGCAGGTC	657							
Oy	1283	CAC TAA G-AGG GTT GAGA AG CTTT GC TC CCAC C TG TGC AG CC	1326							
Db	658	TTTCGAGACAGGTTAAGAGGCTCTGTGACCCCTCGGTTCAGGC	702							
<hr/>										
RESULT 15										
AM572390/c 419 bp mRNA linear EST 13-MAR-2000										
LOCUS xnu09a06.x2 NCI-CGAP_Col4 Homo sapiens cDNA clone IMAGE:2799634 3'										
DEFINITION mRNA sequence.										
ACCESSION AM572390										
VERSION AM572390.1 GI:7237123										
KEYWORDS EST.										
SOURCE Homo sapiens (human)										
ORGANISM Homo sapiens										
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;										
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.										
REFERENCE 1 (bases 1 to 419)										
AUTHORS NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap .										
TITLE National Cancer Institute, Cancer Genome Anatomy Project (CGAP),										
Tumor Gene Index										
JOURNAL Unpublished										
COMMENT Contact: Robert Strausberg, Ph.D.										
Email:rcgapps-re@mail.nih.gov										

Tissue Procurement: Christopher Moskaluk, M.D., Ph.D., Michael R. Emmert-Buck, M.D., Ph.D.

CDNA Library Preparation: Life Technologies, Inc.
CDNA Library Arranged by: Greg Tabor, PhD

CDNA Library Arrayed by: Greg Lennon, Ph.D.
DNA Sequencing by: Washington National Genome Research Institute

DNA sequencing by: Washington University Genome Sequencing Center

Clone distribution: NCI-CCAP clone distribution found through the I.M.A.G.E. Consortium/T.I.N.T. at:

found through the I.M.A.G.E. Consortium/LL
Image.Llnl.gov/Image/html/resources.shtml

Seq primer: -40UP from Glbco

High quality sequence stop: 343.

Location/Qualifiers

1. .419

/organism="Homo sapiens"
/mol_type="mrna"

```
/mol_type="mRNA"  
/db_xref="taxon:9606"
```

```
/dd_xref="taxon:9606"  
/clone="IMAGE:2799634"
```

```
/clone- IMAGE:2/39034
/tissue_type="moderate)
```

```
/lab_host="DH10B"
```

```
clone_11b="NCI_CGAP_Cc
```

```
/note="Organ: colon; Ve
```

Site_2: NotI; Cloned un
Average insert size 1.7

Average Insert size 1.7
11531-019"

11531-019" 122 a 78 c 89 d 1

1 69 100 122

BASE COUNT	122 a	78 c	89 g	130 t
ORIGIN				

ORIGIN

Query Match 16.5%; Score 337.8; DB 9; Length 419;

Best Local Similarity 95.98; Pred. No. 3.3e-78;
Matches 400; Conserved 100; W-amb 100; W-amb 100

Matches	400;	Conservative	0;	Mismatches	12;	Indels	5;	Gaps	5;
---------	------	--------------	----	------------	-----	--------	----	------	----

1630 TCCTGTCCACATCTTGCCCAATGGGAATGGATCTTCCACCAAGAGCTCACCAGCATT 1689

Db 419 TCCCTGTCACATCTTGCCCAATGGGAATGGATCTTTCACCAAGAAGCTCACCAGCATT 360

1690 TTCCACAGAGATGCCAATCTGAGCCCTTGAGGTTCCCAATGGGATTCAGGAAGGAAGT 1749

Db 359 TTCCACAGAGATGCAAATTTGAGCCCTTGAGTTCCAGT-GGATTCAAGGAAGAGT 301

1/50 GGGACAAAGGTGGATGCCCTACTTATGAGCTTGACCAT-ACAGCTATCGGTAATCAGAA 1808

Db 300 GGGACCAAGGTGGATGCCCTACTTATGAGCTTGACCATCACAGCTATCGGTAATCAGAA 241

1809 TATGAACCAAATCTCTG-ACCAAAGAGCAAGCTCTAAGTTCACAAGGTGGCCTT 1867

DB 240 TAATGAACAAATCTCTGCACCAAGAGCAAGCTCTTAAGTTCACAGGGTGCCTGGGCTG 181

1868 GA1011GA1A1ICA-1111CCCT11GCA11111CCCAT1ACATAGAAAC11TGACCCTGTGA 1926

DB 180 CATTTGAAATATCACCCTTCCCCCTCTGCATTTTCCCATCACATAGAGAAGACTTGCACCTGTGAA 121

192/ ACCTGCCAATCTGTATACTAAATTCCCAATAAGGTTCTGTTAGAAATGTCCTTTT 1986

DB 120 GC-TGCCATCTGTAAATACATAAAATGCCCAATAGGTTCTGTTTAGAATGTCCCTTTT 62

198 / AIGCTCTAATATATAGCAGTAAATGTTCAATTTTATGGGATCCIAAAAAAAAA 2043

DB 01 ATGCTTTTAAATATATACGAGTAAATGTTCAATTTTATAGGATCTAAAAA 5

Search completed: August 14, 2003, 17:40:00
Job time : 3872.66 secs

Job time : 3872.66 secs

THIS PAGE BLANK (USPTO)

score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

(without alignments)
11465.210 Million cell updates/sec

Sequence: 1 gaattccattgtgttggtta.....tgggatcctaataaaaaaaa 2043

Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 5777422

```
Minimum DB seq length: 0
Maximum DB seq length: 2000000000
```

Post-processing: Minimum Match 08

Listing first 45 summaries

Database :

```

1:  gb.ba:*
2:  gb.hlg:*
3:  gb_in:*
4:  gb.om:*
5:  gb.ov:*
6:  gb.pat:*
7:  gb.ph:*
8:  gb.pl:*
9:  gb.pr:*
10: gb.ro:*
11: gb.sts:*
12: gb.sy:*
13: gb.un:*
14: gb.v1:*
15: gb.ba:*
16: em.fun:*
17: em_hum:*
18: em_in:*
19: em.om:*
20: em.com:*
21: em.or:*
22: em.ov:*
23: em.pat:*
24: em.ph:*
25: em.pl:*
26: em.ro:*
27: em.sts:*
28: em.un:*
29: em.v1:*
30: em_hlg_hum:*
31: em_hlg_in:*
32: em_hlg_other:*
33: em_hlg_mus:*
34: em_hlg_pln:*
35: em_hlg_rod:*
36: em_hlg_nam:*
37: em_hlg_vtt:*
38: em.sy:*
39: em_hlgo_hum:*
40: em_hlgo_mus:*
41: em_hlgo_other:*

```

Pred. No. is the number of results predicted by chance to have a

Result No.	Score	Query Match	Length	DB	ID	Description
1	1915.2	93.7	2037	9	BC035582	BC035582 Homo sapi
2	1896	92.8	1992	9	AF280088	AF280088 Homo sapi
3	1867.4	91.4	2011	9	AK026635	AK026635 Homo sapi
4	1856.8	90.9	1979	6	BD127258	BD127258 Primer for
5	1856.8	90.9	1979	9	AC074746	AC074746 Homo sapi
6	1856.6	90.9	183288	9	AC010547	AC010547 Homo sapi
7	1856.6	90.9	20593	2	AC138848	AC138848 Homo sapi
8	1813	88.7	2065	6	AR310485	AR310485 Homo sapi
9	1784.8	87.4	2032	6	AR203335	AR203335 Sequence
10	1784.8	87.4	2032	6	BD134772	BD134772 Glycosyls
11	1784.8	87.4	2032	6	AF131335	AF131335 Homo sapi
12	1262	61.8	1333	9	AF149783	AF149783 Homo sapi
13	763.6	37.4	877	6	BD124832	BD124832 Primer for
14	763.6	37.4	877	6	BD126850	BD126850 Primer for
15	755	37.0	2201	10	AF109155	AF109155 Mus muscu
16	752	36.8	232745	2	AC130174	AC130174 Rattus no
17	743.2	36.4	1926	6	BD134773	BD134773 Glycosyls
18	743.2	36.4	1926	10	AF131236	AF131236 Mus muscu
19	708.2	34.7	262679	2	AC110314	AC110314 Rattus no
20	480.2	23.5	517	6	AX381556	AX381556 Sequence
21	412.8	20.2	1647	6	AX327330	AX327330 Sequence
22	412.8	20.2	2254	9	AF219990	AF219990 Homo sapi
23	412.8	20.2	3786	9	AF280086	AF280086 Homo sapi
24	412.8	20.2	71503	9	AF219991	AF219991 Homo sapi
25	412.8	20.2	157358	9	AC009163	AC009163 Homo sapi
26	412.8	20.2	208185	2	AC009105	AC009105 Homo sapi
27	388.4	19.5	1462	9	AF176639	AF176639 Homo sapi
28	388.4	19.5	2110	9	AF176638	AF176638 Homo sapi
29	388.4	19.5	3278	9	AF246718	AF246718 Homo sapi
30	398.4	19.5	194832	9	AC025287	AC025287 Homo sapi
31	390.4	19.1	229007	2	AC095664	AC095664 Rattus no
32	390.4	19.1	245248	2	AC117869	AC117869 Rattus no
33	387.2	19.0	1740	10	AF176841	AF176841 Mus muscu
34	387.2	19.0	1969	10	AF176840	AF176840 Mus muscu
35	387.2	19.0	159170	2	AC115914	AC115914 Mus muscu
36	387.2	19.0	209198	2	AC127315	AC127315 Mus muscu
37	376.2	18.4	548	6	BD125617	BD125617 Primer fo
38	294.4	14.4	389	6	AX381326	AX381326 Sequence
39	144.8	7.1	6961	9	AB017915	AB017915 Homo sapi
40	144.8	7.1	178253	9	AC073370	AC073370 Homo sapi
41	144.8	7.1	196467	9	AC022392	AC022392 Homo sapi
42	141.6	6.9	2156	6	AR071396	AR071396 Sequence
43	141.6	6.9	2156	6	E14937	E14937 Human mRNA
44	141.6	6.9	2731	9	AB012192	AB012192 Homo sapi
45	136.8	6.7	200269	9	AC087442	AC087442 Homo sapi

ALIGNMENTS

RESULT	1
LOCUS	BC035282
DEFINITION	BC035282 2037 bp mRNA linear PRI 23-SEP-2002 Homo sapiens, similar to carbohydrate (N-acetylglucosamine 6-O) sulfotransferase 4, clone MGC:34456 IMAGE:5179826, complete cds.
ACCESSION	BC035282
VERSION	BC035282.1 GI:23273964
KEYWORDS	MGC.
SOURCE	Homo sapiens (human)
ORGANISM	Homo sapiens
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo. 1 (bases 1 to 2037)
AUTHORS	Strausberg,R.

TITLE
JOURNAL

Direct Submission
Submitted (31-Jul-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA

REMARK
COMMENT

NIH-MGC Project URL: <http://mgc.ncl.nih.gov>
Contact: MGC help desk
Email: cgabs-remail.nih.gov
Tissue Procurement: Life Technologies, Inc.
CDNA Library Preparation: Life Technologies, Inc.
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LIML)
DNA Sequencing by: Baylor College of Medicine Human Genome
Sequencing Center
Center code: BCM-HGSC
Web site: <http://www.hgsc.bcm.tmc.edu/cdna/>
Contact: amgdbcm.tmc.edu
Gunaratne, P.H., Garcia, A.M., Lu, X., Huljk, S.W., Hale, S.M.,
Yoon, V.S., Kowis, C.R., Lawrence, S., Martin, R.G., Muzny, D.M.,
Richards, S., Gibbs, R.A.

FEATURES

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LIML at: <http://image.liml.gov>
Series: IRAK Plate: 50 Row: a Column: 7
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 5031734.

source

CDS

Location/Qualifiers
1..2037
/organism="Homo sapiens"
/mol_type="mRNA"
/db_xref="taxon:9606"
/clone="MGC:34456 IMAGE:5179826"
/tissue_type="Brain, Lung, Testis, adult, pooled whole"
/clone_1lb="NH_MGC_115"
/lab_host="DH10B"
/note="Vector: pCMV-SPORT6"
126..1238
/codon_start=1
/product="similar to carbohydrate (N-acetyl)glucosamine
6-O) sulfotransferase 4"
/protein_id="AAH35282.1"
/db_xref="GI:23273965"
/translation="MAILALFEFMYSHNISLSMKRQPERMHLVLSWRSSSFVGO
LEGHDEYILMEPAHWMTFQKSTAWDLIAVAVLCQMSVDAWMEGPR
ROSLFQENSRALCSAPACDIIPODEIPRACHRLCSQPFVEVERKRSYHVL
KEVERFNQSLPLKDPKPSLHIVLHNDPRAVRSRRTGDLIDSRITWQHEO
KAKEDQPYVMQVICOSSOLELYKTQSLPKALQERLYLVREDLARAACTSRYE
FVGLFELPHLQTMVHNITRGKMGMDHAFHTNADALNVSQAWRWSLPYKRVRLQKAC
GDANMLIGYRHVRSEQERNLLDLIDLSLTWVPEQIH"

BASE COUNT

507 a 546 c 481 g 503 t

Query Match

Best Local Similarity 93.7%; Score 1915.2; DB 9; Length 2037;
Matches 1978; Conservative 0; Mismatches 13; Indels 5; Gaps 5;

51 GGGCATGGCCGGGCTAGCAGTGAAGCTCTCAAAAGCAGAGGAAAGCCCAAGCAAG 110
1 GGGCATGGCCGGGCTAGCAGTGAAGCTCTCAAAAGCAGAGGAAAGCCCAAGCAAG 60
111 TCTTCACCTTACGACATGCTACTGCTCTTCAAAATGAAGCTCTGCTTCTGCTT 170
61 TCTTCACCTTACGACATGCTACTGCTCTTCAAAATGAAGCTCTGCTTCTGCTT 120
171 CCCAGATGGCCATCTTGGCTCTATCTTCCACATGATACAGCACAACATCAGCTCCCT 230
121 CCCAGATGGCCATCTTGGCTCTATCTTCCACATGATACAGCACAACATCAGCTCCCT 180
231 CTATGAAGCAGAGCCGAGCGCATGACAGTGTCTGTCTTCTCTGCGCTCTGCGT 290
181 CTATGAAGCAGAGCCGAGCGCATGACAGTGTCTGTCTTCTCTGCGCTCTGCGT 240
291 CTCTCTTGTGGGAGCTTTTGGGAGCAGCAGGAGTCTTCACTGATGAGAGCCG 350
|||||

241 CTCTCTTGTGGGAGCTTTTGGGAGCAGCAGCAGGAGTCTTCACTGATGAGAGCCG 300
351 CCTGGCAGCTGTGATGATGATCTTCAAGCAGAGCAGCGCTGATGCTGACATGCGTGC 410
301 CCTGGCAGCTGTGATGATGATCTTCAAGCAGAGCAGCGCTGATGCTGACATGCGTGC 360
411 GGGATCTGATACGGGCGCTCTTCTGTGCGACATGAGCGCTTTGATGCTCAATGAAC 470
361 GGGATCTGATACGGGCGCTCTTCTGTGCGACATGAGCGCTTTGATGCTCAATGAAC 420
471 CTGGTCCCGGAGAGAGCTGACCGCTTCAAGTGGAGAGAGCGCGCGCTGTCTG 530
421 CTGGTCCCGGAGAGAGCTGACCGCTTCAAGTGGAGAGAGCGCGCGCTGTCTG 480
531 CACCTGCTGTGATCATCATCCACAGATGAATATCCCGGCGCTCACTGAGCTCC 590
481 CACCTGCTGTGATCATCATCCACAGATGAATATCCCGGCGCTCACTGAGCTCC 540
591 TGTGCAATCAACAGCCCTTTGAGTGTGGAGAGAGCGCTGCGCTCTCAAGCAGTGG 650
541 TGTGCAATCAACAGCCCTTTGAGTGTGGAGAGAGCGCTGCGCTCTCAAGCAGTGG 600
651 TGTGCAATCAACAGCCCTTTGAGTGTGGAGAGAGCGCTGCGCTCTCAAGCAGTGG 710
601 TGTGCAATCAACAGCCCTTTGAGTGTGGAGAGAGCGCTGCGCTCTCAAGCAGTGG 660
711 CCTCAACCTGCATATCGTCAACCTGTGCGGAGAGAGCGCGCTGCTCCGAG 770
661 CCTCAACCTGCATATCGTCAACCTGTGCGGAGAGAGCGCGCTGCTCCGAG 720
771 AACGCAAAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 830
721 AACGCAAAAGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
831 AACTCAAG 890
781 AACTCAAG 840
891 AGATCTCAAG 950
841 AGATCTCAAG 900
951 GCTATGAG 1010
1011 GATGGAATCTTGGCCCATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1070
961 GATGGAATCTTGGCCCATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020
1071 TGGGTGACACAGCGCTTTCACACAAATGCCAGGAGATGCCCTTAATGTCTCCAGCTGGC 1130
1021 TGGGTGACACAGCGCTTTCACACAAATGCCAGGAGATGCCCTTAATGTCTCCAGCTGGC 1080
1131 GCTGGCTTGGCCCATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1190
1081 GCTGGCTTGGCCCATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1140
1191 ATTGCTGGGCTACCGCCAGTCAATGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1250
1141 ATTGCTGGGCTACCGCCAGTCAATGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1200
1251 TTCTGTCTACCTGAGCTGTCTGAGCAAAATCCACTAAGAGGTTGAGAGAGCTTTGCTG 1310
1201 TTCTGTCTACCTGAGCTGTCTGAGCAAAATCCACTAAGAGGTTGAGAGAGCTTTGCTG 1260
1311 CCACCTGGGTGAGAGCTGAGTCACTTCTCTGAATGCTTCTGAGAGCTTCTGAGCTTCT 1370
1261 CCACCTGGGTGAGAGCTGAGTCACTTCTCTGAATGCTTCTGAGAGCTTCTGAGCTTCT 1320
1371 GAGCCTTAATCAATGCTGTGGGTATCAACATGAGTGAATGATGATGATGATGATGATG 1430
1321 GAGCCTTAATCAATGCTGTGGGTATCAACATGAGTGAATGATGATGATGATGATGATG 1380

```

QY 1431 AACGAGAGACTTTGTGTCATGCTGTGTCTAGAAAAACAGACTGGGAACTTATGT 1490
    |||
Db 1381 AACGAGAGAGACTTTGTGTCATGCTGTGTCTAGAAAAACAGACTGGGAACTTATGT 1440
QY 1491 GAGCAGACATCCACCACTGAGAAAGGGTATTGCTCTCTCTCTCTCTCTCTCTCT 1550
    |||
Db 1441 GAGCAGACATCCACCACTGAGAAAGGGTATTGCTCTCTCTCTCTCTCTCTCTCT 1500
QY 1551 TCTGGGACACTTCAGAGACTTGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1610
    |||
Db 1501 TCTGGGACACTTCAGAGACTTGTGCTGAGGAGGAGGAGGAGGAGGAGGAGGAG 1560
QY 1611 TGGAAATGATCCATAAACCCTGTCACATCTTCCCAATGGGAATGATCTTTCAC 1670
    |||
Db 1561 TGGAAATGATCCATAAACCCTGTCACATCTTCCCAATGGGAATGATCTTTCAC 1620
QY 1671 CAAGAGCTCACCAGACTTTTCCACAGAGATGCGAAATTCGAGCCCTTGGAGTCCCAAT 1730
    |||
Db 1621 CAAGAGCTCACCAGACTTTTCCACAGAGATGCGAAATTCGAGCCCTTGGAGTCCCAAT 1680
QY 1731 GGGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1789
    |||
Db 1681 -GGATCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1739
QY 1790 AGCTATCGGTAATCAAGAAATATGAAACAAATCTCTG-ACAAAAGAGCAAGCTTAACT 1848
    |||
Db 1740 AGCTATCGGTAATCAAGAAATATGAAACAAATCTCTGACACAAAAGCAAGCTTAACT 1799
QY 1849 TCACAGAGTGGCTGGGCTTGATTTGAATATCA-TTTTCCCTTGCATTTTCCATACATA 1907
    |||
Db 1800 TCACAGAGTGGCTGGGCTTGATTTGAATATCACTTCCCTTGCATTTTCCATACATA 1859
QY 1908 GAAACTTTGACCTGGGAACTTGCATCTGTAATATCTTAATATCTTCCCAATAGGTCT 1967
    |||
Db 1860 GAAACTTTGACCTGGGAACTTGCATCTGTAATATCTTAAATCTTCCCAATAGGTCT 1918
QY 1968 GTTTAGATGTCCTTTTATGCTTAATATTTAGAGTAATATGTCATTTTATGGG 2027
    |||
Db 1919 GTTTAGATGTCCTTTTATGCTTAATATTTAGAGTAATATGTCATTTTATGGG 1978
QY 2028 ATCTTAAAAA 2043
    |||
Db 1979 ATCTTAAAGCTAAAAA 1994

RESULT 2
AF280088. 1992 bp mRNA linear PRI 20-FEB-2001
LOCUS Homo sapiens L-selectin ligand sulfotransferase GST-3 mRNA,
DEFINITION complete cds.
ACCESSION AF280088
VERSION AF280088.1 GI:12060807
KEYWORDS Homo sapiens (human)
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 1992)
AUTHORS Hemmerich,S., Lee,J.K., Bhakta,S., Bistrup,A., Ruddle,N.R. and
Rosen,S.D.
TITLE Chromosomal localization and genomic organization for the
JOURNAL galactose/N-acetylgalactosamine/N-acetylglucosamine
MEDLINE 6-O-sulfotransferase gene family
PUBMED glycolysis 11 (1), 75-87 (2001)
21096027
11181564
REFERENCE 2 (bases 1 to 1992)
AUTHORS Hemmerich,S., Bhakta,S., Lee,J.-K., Bistrup,A., Ruddle,N.R. and
Rosen,S.D.
TITLE Direct Submission
JOURNAL Submitted (20-JUN-2000) Respiratory Diseases, Roche Bioscience,
3401 Hillview Ave., Palo Alto, CA 94304, USA

```

```

FEATURES
    source
        Location/Qualifiers
            1..1992
                /organism="Homo sapiens"
                /mol_type="mRNA"
                /db_xref="taxon:9606"
                /chromosome="16"
                /map="16q23.1-q23.2"
                /cell_type="high endothelial"
                /tissue_type="tonsil"
            1..1992
                /gene="GST3"
            122..1282
                /gene="GST3"
            CDS
                /note="similar to Homo sapiens sequence encoded by GenBank
                Accession Number AF131235"
                /codon_start=1
                /product="L-selectin ligand sulfotransferase GST-3"
                /protein_id="AAG48246.1"
                /db_xref="GI:12060808"
                /translation="MLPKMKLLFLVSOMAILLFFHMYSHNISLSMKQPERH
                VLVSSMRSSSFVGLFGHPDVFYLMPEAMHVMTPFQSTAMLMAMVRODLRAVF
                LCDMSYFDAMMEPRROSLFPMENSRLCSRPACDIIDPDIIRAPRRLCSOP
                FEVEKACRSYSHVLEKVEFNQLSLPLKPSLNLHLVLRDPRALRERERK
                GDLMIDSRIMGHEQKLEKEDQPYVMQVYTCQSOLEIYTIQSLPRALQERYLVAY
                EDLARPVQTSRYEFVGLFPLHQTWVHNITRKGMDHAFHTNARDALNVSOAM
                RMSLPYKYSRLQKACGDAMNLLGSTRHVRSEQGRNLLDLSTWTVPEQIH"
BASE COUNT      482 a      540 c      489 g      481 t
ORIGIN
Query Match      92.8%; Score 1896; DB 9; Length 1992;
Best Local Similarity 99.2%; Pred. No. 0;
Matches 1957; Conservative 0; Mismatches 10; Indels 5; Gaps 5;
QY 7 CATTGTGTGGGTACGAGAGACGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 66
    |||
Db 1 CATTGTGTGGGTACGAGAGACGACGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 60
QY 67 GCAGTGAAGCTTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 126
    |||
Db 61 GCAGTGAAGCTTCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120
QY 127 AATGCTACTGCTTAAAAAATGAGAGCTGCTGCTTGTGTTTCCAGATGGCATCTT 186
    |||
Db 121 AATGCTACTGCTTAAAAAATGAGAGCTGCTGCTTGTGTTTCCAGATGGCATCTT 180
QY 187 GGCCTATTTCTTCCACATGTACAGCCACCAATCAGTCCCTGTCTATGAGGACAGGCC 246
    |||
Db 181 GGCCTATTTCTTCCACATGTACAGCCACCAATCAGTCCCTGTCTATGAGGACAGGCC 240
QY 247 CGAGCCCATGACAGTGTGCTGCTTCTGCTTCTGCGGCTGCGCTCTTTTGTGGGCA 306
    |||
Db 241 CGAGCCCATGACAGTGTGCTGCTTCTGCTTCTGCGGCTGCGCTCTTTTGTGGGCA 300
QY 307 GCTTTTGGGACACCCAGATGTTTCTACATGATGAGACCCGCTGAGCATGTGAT 366
    |||
Db 301 GCTTTTGGGACACCCAGATGTTTCTACATGATGAGACCCGCTGAGCATGTGAT 360
QY 367 GACCTTCACAGACAGACCCGCTGAGATGTCACATGAGCTGCGGAGATGTATAGGGC 426
    |||
Db 361 GACCTTCACAGACAGACCCGCTGAGATGTCACATGAGCTGCGGAGATGTATAGGGC 420
QY 427 CGTCTTCTGTGAGCATGAGCTGTTTGAATGCTTACATGAACTGTGCTCCGAGACA 486
    |||
Db 421 CGTCTTCTGTGAGCATGAGCTGTTTGAATGCTTACATGAACTGTGCTCCGAGACA 480
QY 487 GTCCAGCTCTTTCAGTGGAGAGACCGGGCCCTGTGTGTCACCTGTCATCAT 546
    |||
Db 481 GTCCAGCTCTTTCAGTGGAGAGACCGGGCCCTGTGTGTCACCTGTCATCAT 540
QY 547 CATCCCAAGAGAGAAATCATCCCGGGGCTCATGAGGCTCTGTCAGCATACAGCC 606
    |||
Db 541 CATCCCAAGAGAGAAATCATCCCGGGGCTCATGAGGCTCTGTCAGCATACAGCC 600

```

Db	QY	1747	AGTGGGAACAAGTTGGATGGCTACTTATGAGCTTGACAT-ACAGTATTCGTAATCAG	1805
Db	1681	ATTTCACAGAGATGCGAATTCGACCCCTTGAGTTCACCAGT-GGATTCACAGAAGAA	1739	
QY	1740	AGTGGGAACAAGTTGGATGGCTACTTATGAGCTTGACATCAGAGTATTCGTAATCAG	1799	
Db	1806	AAATATGAACAAATATCTCTG-ACAAAAGCAAGCTTAAGTTCACAAAGTCCCTGGG	1864	
QY	1800	AAATATGAACAAATATCTCTG-ACAAAAGCAAGCTTAAGTTCACAAAGTCCCTGGG	1859	
Db	1865	CTGATATGGAATATCA-TTTCCTCTGTCATTTTCCATTCATAGAAACCTTGACCTGT	1923	
QY	1860	CTGATATGGAATATCA-TTTCCTCTGTCATTTTCCATTCATAGAAACCTTGACCTGT	1919	
Db	1924	GAAACCTGCACTCTGTTAATATCTAAATTCCCAAATTAAGTTCTGTTAGAA	1975	
QY	1920	GAAACCTGCACTCTGTTAATATCTAAATTCCCAAATTAAGTTCTGTTAGAA	1970	
RESULT 3	AK026635	2011 bp	linear	PRI 29-SEP-2000
LOCUS	AK026635			
DEFINITION	Homo sapiens cDNA: FLJ22982 fis, clone KAT11454, highly similar to AF131335 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA.			
ACCESSION	AK026635	GI:10439531		
VERSION	AK026635.1			
KEYWORDS	oligo capping; fis (full insert sequence).			
SOURCE	Homo sapiens (human)			
ORGANISM	Homo sapiens			
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.			
AUTHORS	1 (sites)			
TITLE	Watanabe,K., Kumagai,A., Itakura,S., Yamazaki,M., Tashiro,H., Ota,T., Suzuki,Y., Obayashi,M., Nishi,T., Shibahara,T., Tanaka,T., Nakamura,Y., Isogai,T. and Sugano,S.			
JOURNAL	NEDO human cDNA sequencing project			
REFERENCE	Unpublished			
AUTHORS	2 (bases 1 to 2011)			
TITLE	Sugano,S., Suzuki,Y., Ota,T., Obayashi,M., Nishi,T., Isogai,T., Shibahara,T., Tanaka,T. and Nakamura,Y.			
JOURNAL	Direct Submission			
COMMENT	Submitted (29-AUG-2000) Sumio Sugano, Institute of Medical Science, University of Tokyo, Laboratory of Genome Structure Analysis, Human Genome Center, Shirokane-dai, 4-6-1, Minato-ku, Tokyo 108-8639, Japan (E-mail:cdna@ims.u-tokyo.ac.jp, Tel:81-3-5449-5286, Fax:81-3-5449-5416)			
FEATURES	NEDO human cDNA sequencing project supported by Ministry of International Trade and Industry of Japan: cDNA full insert sequencing: Research Association for Biotechnology: cDNA library construction, 5'- & 3'-end one pass sequencing: Department of Virology and Human Genome Center, Institute of Medical Science, University of Tokyo (partly supported by Science and Technology Agency).			
SOURCE	Location/Qualifiers			
FEATURES	1..2011			
FEATURES	/organism="Homo sapiens"			
FEATURES	/mol_type="mRNA"			
FEATURES	/db_xref="taxon:9606"			
FEATURES	/clone="KAT11454"			
FEATURES	/cell_line="KATO III"			
FEATURES	/cell_type="signal-ring cell carcinoma"			
FEATURES	/clone_id="KAT"			
FEATURES	/note="Cloning vector PME18SFL3"			
FEATURES	1..2011			
FEATURES	/note="highly similar to AF131235 Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA"			
BASE COUNT	482 a	541 c	475 g	513 t
ORIGIN				
Query Match	91.4%	Score 1867.4	DB 9	Length 2011
Best Local Similarity	98.9%	Pred. No. 0		

Matches 1932; Conservative 0; Mismatches 16; Indels 5; Gaps 5;

```

OY 94 AAGCCAAAGCCAAAGGCTTCCACTTCCAGACATGCTACTGCTTAAAAAATTAACCT 153
    |||||
Db 38 AAGCCCGCTTGTGAAAGCTTTCACCTTACGACAAATGCTACTGCTTAAAAAATTAACCT 97
OY 154 CCTCTGTTTGTGTTTCCAGATGAGCCATTTGGCTTATTTCTCCACATGTACAGCA 213
    |||||
Db 98 CCTCTGTTTGTGTTTCCAGATGAGCCATTTGGCTTATTTCTCCACATGTACAGCA 157
OY 214 CAACATCAGCTCCCTGTCTATGAAGGACACGCCGAGCGCATGCATGCTGTTCTGTC 273
    |||||
Db 158 CAACATCAGCTCCCTGTCTATGAAGGACACGCCGAGCGCATGCATGCTGTTCTGTC 217
OY 274 TTCTGGGCGCTGCTGCTTCTTTTGTGGGGGAGCTTTTGGGAGCCCGCAAGATCTT 333
    |||||
Db 218 TTCTGGGCGCTGCTGCTTCTTTTGTGGGGGAGCTTTTGGGAGCCCGCAAGATCTT 277
OY 334 CTACCTGATGAGCCCGCTGCGACGCTGTGATGACCTTCAAGCAGACACCGCTGAT 393
    |||||
Db 278 CTACCTGATGAGCCCGCTGCGACGCTGTGATGACCTTCAAGCAGACACCGCTGAT 337
OY 394 GCTGCAATGCTGCTGTGGGAGATCTGATACGGGCGCTCTTGTGCGCATGAGCGCTT 453
    |||||
Db 338 GCTGCAATGCTGCTGTGGGAGATCTGATACGGGCGCTCTTGTGCGCATGAGCGCTT 397
OY 454 TGATGCTACATGGAACCTGCTCCCGGAGACGTCGCGCTTTTGTGCGGAGAAAG 513
    |||||
Db 398 TGATGCTACATGGAACCTGCTCCCGGAGACGTCGCGCTTTTGTGCGGAGAAAG 457
OY 514 CCGGGCCCTGTGTTCTGACACCTGCTGTGATGACATACCCACAGATGAATCATCCCG 573
    |||||
Db 458 CCGGGCCCTGTGTTCTGACACCTGCTGTGATGACATACCCACAGATGAATCATCCCG 517
OY 574 GGCTCATGCGAGGCTCCTGTGACATCAACACGCCCTTTAGTGTGGGAAAGGCTGCG 633
    |||||
Db 518 GGCTCATGCGAGGCTCCTGTGACATCAACACGCCCTTTAGTGTGGGAAAGGCTGCG 577
OY 634 CTCTACAGCCAGCTGTGCTCAAGAGAGTGCGCTTCTTCAACCTGAGTCCCTTACC 693
    |||||
Db 578 CTCTACAGCCAGCTGTGCTCAAGAGAGTGCGCTTCTTCAACCTGAGTCCCTTACC 637
OY 694 GCTGCTGAAGACCCCTCCCTCAACCTGATATGCTGACCTGTGTCGCGGAGCCCGGG 753
    |||||
Db 638 GCTGCTGAAGACCCCTCCCTCAACCTGATATGCTGACCTGTGTCGCGGAGCCCGGG 697
OY 754 CGTGTGCGTCCCGAGAGGACAAAGGAGATCTCATGATGACATCCCATTTGAT 813
    |||||
Db 698 CGTGTGCGTCCCGAGAGGACAAAGGAGATCTCATGATGACATCCCATTTGAT 757
OY 814 GGGGAGCATAGGAAAAAATCAAGAGAGAGGACCAACCTACTATGATGAGAGGTCAT 873
    |||||
Db 758 GGGGAGCATAGGAAAAAATCAAGAGAGAGGACCAACCTACTATGATGAGAGGTCAT 817
OY 874 CTGCCAAAGCCAGCTGAGATCTACAGACCATCAGTCTTGGCCCAAGGCCCTGACAGA 933
    |||||
Db 818 CTGCCAAAGCCAGCTGAGATCTACAGACCATCAGTCTTGGCCCAAGGCCCTGACAGA 877
OY 934 ACGTACCTGCTTGTGCGCTATGAGAGACCTGGCTGAGACCCCTGTGGCCAGACTCCCG 993
    |||||
Db 878 ACGTACCTGCTTGTGCGCTATGAGAGACCTGGCTGAGACCCCTGTGGCCAGACTCCCG 937
OY 994 AATGTATGAATTCGTGGATGGAATTCCTGCCCATTTAGACCTGGTGCATTAACAT 1053
    |||||
Db 938 AATGTATGAATTCGTGGATGGAATTCCTGCCCATTTAGACCTGGTGCATTAACAT 997
OY 1054 CACCCGAGGCAAGGCGATGAGTGAACACGCTTTCACACAATGCGCAGGATGCCCTTAA 1113
    |||||
Db 998 CACCCGAGGCAAGGCGATGAGTGAACACGCTTTCACACAATGCGCAGGATGCCCTTAA 1057
OY 1114 TGCTCCAGGCTGTGGGCGTGTGTTGCTTATGAAGAAAGTTTCTGCATCTCAGAAAGC 1173
    |||||
Db 1058 TGCTCCAGGCTGTGGGCGTGTGTTGCTTATGAAGAAAGTTTCTGCATCTCAGAAAGC 1117

```

```

OY 1174 CTGTGGCATGCGATGATTTGTGCGGCTACCGCCAGCGCATGCTGAACAAGACAGAG 1233
    |||||
Db 1118 CTGTGGCATGCGATGATTTGTGCGGCTACCGCCAGCGCATGCTGAACAAGACAGAG 1177
OY 1234 AAACCTGTGTGATCTTCTGTCTACCTGAGACTGCTCCCTGAGCAAAATCCATAAGAGG 1293
    |||||
Db 1178 AAACCTGTGTGATCTTCTGTCTACCTGAGACTGCTCCCTGAGCAAAATCCATAAGAGG 1237
OY 1294 TTGAGAGGCTTGTGCTGCGACCTGTGTGACCTTCAGTCACTTCTCTGAATGCTTCTGA 1353
    |||||
Db 1228 TTGAGAGGCTTGTGCTGCGACCTGTGTGACCTTCAGTCACTTCTCTGAATGCTTCTGA 1297
OY 1354 GCGTTGCGTACATCTGAGCCTTACATACATGCTGTGGGATATACACTGAGTGTGAGT 1413
    |||||
Db 1298 GCGTTGCGTACATCTGAGCCTTACATACATGCTGTGGGATATACACTGAGTGTGAGT 1357
OY 1414 TGTGTCCACAGTGTGCTCAAGCAGAAAGACTTTTGTCCATGCTGTGTCTAGAAAAGAG 1473
    |||||
Db 1358 TGTGTCCACAGTGTGCTCAAGCAGAAAGACTTTTGTCCATGCTGTGTCTAGAAAAGAG 1417
OY 1474 ACTGGGAACTTATATGAGACGACATCCACAGTGAACAGGATATGCTCTTCTTC 1533
    |||||
Db 1418 ACTGGGAACTTATATGAGACGACATCCACAGTGAACAGGATATGCTCTTCTTC 1477
OY 1534 TTTTCTGATCTCTCTGCTGGGAGACTTTCAGACATTTTGTGGCTGGAGGCTTATTA 1593
    |||||
Db 1478 TTTTCTGATCTCTCTGCTGGGAGACTTTCAGACATTTTGTGGCTGGAGGCTTATTA 1537
OY 1594 GCACGACACAGTATACGTAATGATATCATTAACCTCCCTGCTCCACATCTTGCCCAATG 1653
    |||||
Db 1538 GCACGACACAGTATACGTAATGATATCATTAACCTCCCTGCTCCACATCTTGCCCAATG 1597
OY 1654 GGGATGATCTTTTACCAAGAGCTCACACAGATTTTCCACAGAGATGCCAATCTGAG 1713
    |||||
Db 1598 GGGATGATCTTTTACCAAGAGCTCACACAGATTTTCCACAGAGATGCCAATCTGAG 1657
OY 1714 CCGTTGAGTTCCTCAATGGATTCAGAGAGAGTGGGAAACAAAGTGGATGCTACTT 1773
    |||||
Db 1658 CCGTTGAGTTCCTCAATGGATTCAGAGAGAGTGGGAAACAAAGTGGATGCTACTT 1716
OY 1774 ATGACCTTGACCAT-ACAGCTATGCTGTAATCAGAAATATGAACAAATCTCTG-ACAAA 1831
    |||||
Db 1717 ATGACCTTGACCATACAGCTATGCTGTAATCAGAAATATGAACAAATCTCTGACAAA 1776
OY 1832 AGAGCAAGCTCTTAAGTACAGAGTGCTGGGCTGTAATGAATATCA-TTTCCCTTGG 1890
    |||||
Db 1777 AGAGCAAGCTCTTAAGTACAGAGTGCTGGGCTGCAATGAATATCACTTCCCTCTG 1836
OY 1891 CATTTCCATTTACATAAAGAACTTTGACCTGTGAACCTTGCCATCTGTAATATAAAA 1950
    |||||
Db 1837 CATTTCCATTTACATAAAGAACTTTGACCTGTGAACCTTGCCATCTGTAATATAAAA 1895
OY 2011 ATGCTCATTTTATGCGATCCCTAAAAAATTA 2043
    |||||
Db 1956 ATGCTCATTTTATGCGATCCCTAAAAAATTA 1988

```

RESULT 4
 LOCUS BD127258 1979 bp DNA linear PAT 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD127258
 VERSION BD127258.1 GI:23222203
 KEYWORDS JP 2002017375-A/2689.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE 1 (bases 1 to 1979)
 AUTHORS Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
 Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
 Koga,H.
 TITLE Primer for synthesizing full-length cDNA and use thereof
 JOURNAL Patent: JP 2002017375-A 2689 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 COMMENT OS Homo sapiens (human)
 PN JP 2002017375-A/2689
 PD 22-JAN-2002
 PE 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA, TETSUO NISHIKAWA, TAKAO ISOGAI, KOJI HAYASHI, SHIZUKO
 PI ISHII
 PI YUKI KAWAI, AI WAKAMATSU, TOMOYASU SUGIYAMA, KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI, HISASHI KOGA
 PC
 C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10', C12P21/02,C12Q1/68//C12P21/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH key
 Location/Qualifiers
 FT CDS
 Location/Qualifiers
 1..1979
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 BASE COUNT 451 a 542 c 474 g 512 t
 ORIGIN
 Query Match 90.98; Score 1856.8; DB 6; Length 1979;
 Best Local Similarity 98.98; Pred. No. 0;
 Matches 1922; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

Db 575 CTCCTACAGCCAGCTGCTCTCAAGAGAGTGCCCTTCTTCAACCTGCAATCCCTACACC 634
 QY 694 GCTGCTGAAGAACCCCTCCCTCAACCTGCATATGTCACCTGGTCCGGAGCCCGGGC 753
 Db 635 GCGTCTGAAGAACCCCTCCCTCAACCTGCATATGTCACCTGGTCCGGAGCCCGGGC 694
 QY 754 CGTGTCCGTTCCCGAGAACGCAACAAAGGAGATCTCATGATTGACAGTCCGATTGTGAT 813
 Db 695 CGTGTCCGTTCCCGAGAACGCAACAAAGGAGATCTCATGATTGACAGTCCGATTGTGAT 754
 QY 814 GGGGCGACATGACCAAAAACTCAAGAAAGAGAGACCAACCTACTATCTGATGACAGTCAT 873
 Db 755 GGGGCGACATGACCAAAAACTCAAGAAAGAGAGACCAACCTACTATGATGACAGTCAT 814
 QY 874 CTCGCAAGCCACCTGAGATCTACAAGACCATTCAGTCTTGCCCAAGGCCCTGCAGGA 933
 Db 815 CTCGCAAGCCACCTGAGATCTACAAGACCATTCAGTCTTGCCCAAGGCCCTGCAGGA 874
 QY 934 AGCTACCTGCTTGTGCGCTATAGAGACCTGGCTCGAGCCCTGTGGCCAGACTTCCG 993
 Db 875 AGCTACCTGCTTGTGCGCTATAGAGACCTGGCTCGAGCCCTGTGGCCAGACTTCCG 934
 QY 994 AATGTATGATTTGCTGGGATTTGGAATTTCTGCCCCATCTTCACACCTGGGTGATTAAT 1053
 Db 935 AATGTATGATTTGCTGGGATTTGGAATTTCTGCCCCATCTTCACACCTGGGTGATTAAT 994
 QY 1054 CACCCGAGCAAGGAGATGGGTGACCAACGCTTTCACACAAATGCCAGGATCCCTTAA 1113
 Db 995 CACCCGAGCAAGGAGATGGGTGACCAACGCTTTCACACAAATGCCAGGATCCCTTAA 1054
 QY 1114 TGTCTCCAGGCTTGGCGCTGTGCTTGGCCATATGAAAAGTTTCTGACTGAGAAAGC 1173
 Db 1055 TGTCTCCAGGCTTGGCGCTGTGCTTGGCCATATGAAAAGTTTCTGACTGAGAAAGC 1114
 QY 1174 CTGTGGGAGATGACATGATTTGCTGGGCTACCGCCACGCTGACATGGAACAAAGAGAG 1233
 Db 1115 CTGTGGGAGATGACATGATTTGCTGGGCTACCGCCACGCTGACATGGAACAAAGAGAG 1174
 QY 1234 AAACCTGTCCTGATCTTCTGTCTACCTGAGATGTCCTCGAGCAAAATCCATTAAGAGG 1293
 Db 1175 AAACCTGTCCTGATCTTCTGTCTACCTGAGATGTCCTCGAGCAAAATCCATTAAGAGG 1234
 QY 1294 TTGAGAAAGCTTGTGCTGACCACTGTGTGACGCTCACTGCTTCTGTGAACTGTGA 1353
 Db 1235 TTGAGAAAGCTTGTGCTGACCACTGTGTGACGCTCACTGCTTCTGTGAACTGTGA 1294
 QY 1354 GCCTTGCTACATCTCTGAGCCCTTAATCATGTCGTGGGCTTCAACATGAGTGTGAGT 1413
 Db 1295 GCCTTGCTACATCTCTGAGCCCTTAATCATGTCGTGGGCTTCAACATGAGTGTGAGT 1354
 QY 1414 TGTGTCCACAGCTGCTCAAGCAGAAAGACTTTTGTGTCAATGCTTGTCTAGAAAACAG 1473
 Db 1355 TGTGTCCACAGCTGCTCAAGCAGAAAGACTTTTGTGTCAATGCTTGTCTAGAAAACAG 1414
 QY 1474 ACTGGGGAACCTTATGTGAGCAGACATATCCACAGTGAAGAGGATATTGCTTCTTTC 1533
 Db 1415 ACTGGGGAACCTTATGTGAGCAGACATATCCACAGTGAAGAGGATATTGCTTCTTTC 1474
 QY 1534 TTTTCTTGAATCTTCTGCTGGGAGAGCTTCAAGACTTGTGGCTGGAGGCTTATTA 1593
 Db 1475 TTTTCTTGAATCTTCTGCTGGGAGAGCTTCAAGACTTGTGGCTGGAGGCTTATTA 1534
 QY 1594 GCAAGCAGATATCAGTGAATTTGATTCATAAATCCCTGTCACATCTTGCCCAATG 1653
 Db 1535 GCAAGCAGATATCAGTGAATTTGATTCATAAATCCCTGTCACATCTTGCCCAATG 1594
 QY 1654 GGAATGATCTTTCACCAAGAGCTCAGCAGATTTTCCACAGAGATGCGAATTTGAG 1713
 Db 1595 GGAATGATCTTTCACCAAGAGCTCAGCAGATTTTCCACAGAGATGCGAATTTGAG 1654
 QY 1714 CCCTTGAGTTCACATGAGATTTCAAGGAAGAGGGAACAAGTTGATGAGCCACTT 1773
 Db 1655 CCCTTGAGTTCACATGAGATTTCAAGGAAGAGGGAACAAGTTGATGAGCCACTT 1713

QY 994 AATGATGAATTCGTGGGATGGAATTCCTGCCCCATCTTCAGACCTGGGTCATAACAT 1053
 DB 935 AATGATGAATTCGTGGGATGGAATTCCTGCCCCATCTTCAGACCTGGGTCATAACAT 994
 QY 1054 CACCCGAGGAGGATGGGTCACACAGCTTCCACACAAATGCCAGGATGCCCTTAA 1113
 DB 995 CACCCGAGGAGGATGGGTCACACAGCTTCCACACAAATGCCAGGATGCCCTTAA 1054
 QY 1114 TGTCTCCAGAGCTGGGTCGCTGCTTGGCCCTATGAAAAGTTTCTGAGTGAAGAGC 1173
 DB 1055 TGTCTCCAGAGCTGGGTCGCTGCTTGGCCCTATGAAAAGTTTCTGAGTGAAGAGC 1114
 QY 1174 CTGTGGCGATGCCATGAATTCCTGGGCTACCCGACGTCAGATCTGACACAGACAG 1233
 DB 1115 CTGTGGCGATGCCATGAATTCCTGGGCTACCCGACGTCAGATCTGACACAGACAG 1174
 QY 1234 AATCCTGTCGATCTTCTGTCTACCTGAGCTGACCTGACCAATCCATGAAGGG 1293
 DB 1175 AATCCTGTCGATCTTCTGTCTACCTGAGCTGACCTGACCAATCCATGAAGGG 1234
 QY 1294 TTGAGAGGCTTGTGTCACACCTGCTGTCAGCTCAGCTTCTCTGTAATGCTCTGA 1353
 DB 1235 TTGAGAGGCTTGTGTCACACCTGCTGTCAGCTCAGCTTCTCTGTAATGCTCTGA 1294
 QY 1354 GCCTTGCCTACATCTCTGAGCCCTTAATGATGTCGTGGGTATCACACTGAGTGTAGT 1413
 DB 1295 GCCTTGCCTACATCTCTGAGCCCTTAATGATGTCGTGGGTATCACACTGAGTGTAGT 1354
 QY 1414 TGTGTCCACAGCTGTCTCAGACAGAGGACTTGTGTGCATGCTGTGTAGAAAACAG 1473
 DB 1355 TGTGTCCACAGCTGTCTCAGACAGAGGACTTGTGTGCATGCTGTGTAGAAAACAG 1414
 QY 1474 ACTGGGGAACCTTATGTAGACACACATCCCAAGGATGATGCTCTCTTC 1533
 DB 1415 ACTGGGGAACCTTATGTAGACACACATCCCAAGGATGATGCTCTCTTC 1474
 QY 1534 TTTTCTGATCTCTGCTGGGAGCTTCAGAGCTTGGCCCTGGAGGCTCTTAA 1593
 DB 1475 TTTTCTGATCTCTGCTGGGAGCTTCAGAGCTTGGCCCTGGAGGCTCTTAA 1534
 QY 1594 GCACGACACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTCCCAATG 1653
 DB 1535 GCACGACACAGTATCAGTGAATTCATTAACCTCCCTGTCACATCTTCCCAATG 1594
 QY 1654 GGAATGAGATCTTTCACAAAGGCTCAGAGATTTTCCACAGAGATGCAATTCGAG 1713
 DB 1595 GGAATGAGATCTTTCACAAAGGCTCAGAGATTTTCCACAGAGATGCAATTCGAG 1654
 QY 1714 CCCTTGGAGTCCCAATGGATTCAGGAAGTGGGAACAAGTTGATGCTCTACT 1773
 DB 1655 CCCTTGGAGTCCCAATGGATTCAGGAAGTGGGAACAAGTTGATGCTCTACT 1713
 QY 1774 ATGAGCTTGACAT-ACAGCTATCGGATTCAGAAAATGAAAACAAATCTCTG-ACAAA 1831
 DB 1714 ATGAGCTTGACATCAGAGCTATCGGATTCAGAAAATGAAAACAAATCTCTGACAAA 1773
 QY 1832 AGGCAAGCTCTTAATTCACAGAGTCCCTGGGCTGATGTAATCA-TTTCCTTTG 1890
 DB 1774 AGGCAAGCTCTTAATTCACAGAGTCCCTGGGCTGATGTAATCA-TTTCCTTTG 1833
 QY 1891 CATTTTCCATACATAGAAAATTTGACCTGGAACCTTGATCGATCGTAATACATAAA 1950
 DB 1834 CATTTTCCATACATAGAAAATTTGACCTGGAACCTTGATCGATCGTAATACATAAA 1892
 QY 1951 TTCCCAATAGAGTCTGTTTGAATGTCCTTTTATGCTCTTATTTATAGAGTAA 2010
 DB 1893 TTCCCAATAGAGTCTGTTTGAATGTCCTTTTATGCTCTTATTTATAGAGTAA 1952
 QY 2011 ATGTTTATTTTATGGGATCTTAA 2034
 DB 1953 ATGTTTATTTTATGGGATCTTAA 1976

RESULT 6
 AC010547
 LOCUS 183228 bp DNA linear PRI 29-SEP-2001
 DEFINITION Homo sapiens chromosome 16 clone RP11-510M2, complete sequence.
 AC010547
 VERSION AC010547.9 GI:15808510
 KEYWORDS HTG.
 SOURCE
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE DOE Joint Genome Institute.
 JOURNAL Unpublished
 REFERENCE
 AUTHORS DOE Joint Genome Institute.
 TITLE DOE Joint Genome Institute.
 JOURNAL Direct Submission
 REFERENCE
 AUTHORS Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 TITLE DOE Joint Genome Institute.
 JOURNAL Direct Submission
 REFERENCE
 AUTHORS Submitted (29-SEP-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
 TITLE On Sep 29, 2001 this sequence version replaced g1:14589436.
 COMMENT
 Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

 Sequence Quality Assessment:
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
 All manually edited bases have been reduced to quality zero.
 Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
 Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.

 FEATURES
 source
 1. 183228
 /organism="Homo sapiens"
 /mol_type="genomic DNA"
 /db_xref="taxon:9606"
 /chromosome="16"
 /clone="RP11-510M2"
 /library="RP11-510M2"
 BASE COUNT 51901 a 42098 c 40026 g 49203 t
 ORIGIN
 Query Match 90.9%; Score 1856.6; DB 9; Length 183228;
 Best Local Similarity 99.0%; Pred. No. 0;
 Matches 1920; Conservative 0; Mismatches 14; Indels 5; Gaps 5;
 QY 107 AAGGTCTCCACTTCAGACAAATGCTACGTCCTTAAATAAATGAAGTCCGCTGTTTCG 166
 DB 142584 AAGGTCTCCACTTCAGACAAATGCTACGTCCTTAAATAAATGAAGTCCGCTGTTTCG 142643
 QY 167 GTTCCAGATGAGCATCTTGGCTCTATTCTTCACATGTACAGCCACAATCAGCTCC 226
 DB 142644 GTTCCAGATGAGCATCTTGGCTCTATTCTTCACATGTACAGCCACAATCAGCTCC 142703
 QY 227 CTGTCTATGAGGACAGCCGAGCGCATGACGTCGTGTTCTGTTCTGCGCTCT 286
 DB 142704 CTGTCTATGAGGACAGCCGAGCGCATGACGTCGTGTTCTGTTCTGCGCTCT 142763
 QY 287 GGCTCTCTTTTGTGGGAGCTTTTGGGACAGCACCAGATGTTTCTACCTGATGAG 346

```
|||||
Db 142764 GGGCTCTCTTTGGGGGAGCTTTTGGGAGACCCAGATGTTTCTACCTGATGAG 142823
QY 347 CCGGCTGGGACGCTGGGATGACCTTCAACAGAGCACCGCTGGATGCTGACATGCT 406
Db 142824 CCGGCTGGGACGCTGGGATGACCTTCAACAGAGCACCGCTGGATGCTGACATGCT 142883
QY 407 GTCGGGATCTGATACGGGCGCTTCTTGGGACATGAGCGCTTGTGATGCTTACATG 466
Db 142884 GTGGGGATCTGATACGGGCGCTTCTTGGGACATGAGCGCTTGTGATGCTTACATG 142943
QY 467 GAACCTGGTCCCGGAGACAGTCCAGGCTTTTCAAGTGGAGAACACCGGCGCTGTGT 526
Db 142944 GAACCTGGTCCCGGAGACAGTCCAGGCTTTTCAAGTGGAGAACACCGGCGCTGTGT 143003
QY 527 TCGGACCTGCTGTGATCATATCCCAAGATGAAATCATCCCGGCGCTGCTACAGCAC 586
Db 143004 TCGGACCTGCTGTGATCATATCCCAAGATGAAATCATCCCGGCGCTGCTACAGCAC 143063
QY 587 CTCTGTGACAGTCAACAGCCCTTGTGAGTGGTGGAGAGGCTCCGCTCTACAGCAC 646
Db 143064 CTCTGTGACAGTCAACAGCCCTTGTGAGTGGTGGAGAGGCTCCGCTCTACAGCAC 143123
QY 647 GTGGTGTCAAGAGGTGGCGCTTCTTCAACCTGCAAGTCCCTTACCCGCTGTGAAAGAC 706
Db 143124 GTGGTGTCAAGAGGTGGCGCTTCTTCAACCTGCAAGTCCCTTACCCGCTGTGAAAGAC 143183
QY 707 CCTCTCCCTCAACCTGCAATATGTCGACACTGTCCGGGACCCCGGCGCTGTCCGCTTCC 766
Db 143184 CCTCTCCCTCAACCTGCAATATGTCGACACTGTCCGGGACCCCGGCGCTGTCCGCTTCC 143243
QY 767 CGAGAACGCAACAAAGGAGATCTCATGATTGACAGTGGCATTTGATGGGGACAGATGAG 826
Db 143244 CGAGAACGCAACAAAGGAGATCTCATGATTGACAGTGGCATTTGATGGGGACAGATGAG 143303
QY 827 CAAAACTCAAGAGAGAGCAACCTACTATGTGATGAGATCTGTGCAAAAGCAG 886
Db 143304 CAAAACTCAAGAGAGAGCAACCTACTATGTGATGAGATCTGTGCAAAAGCAG 143363
QY 887 CTGAGATCTTCAAGACATCTGACACTTGTGCGCAAGGCTGTGCAAGAGCTTACTGCTT 946
Db 143364 CTGAGATCTTCAAGACATCTGACACTTGTGCGCAAGGCTGTGCAAGAGCTTACTGCTT 143423
QY 947 GTGGCTATGAGAGCTGTGCTGAGGCGCTGTGCGGAGAGCTTCCGCAATGTATGAATTC 1006
Db 143424 GTGGCTATGAGAGCTGTGCTGAGGCGCTGTGCGGAGAGCTTCCGCAATGTATGAATTC 143483
QY 1007 GTGGGATTTGGAATTTCTTCCCATCTTTCAGACCTGGGTGATTAACATCACCCGAGCAG 1066
Db 143484 GTGGGATTTGGAATTTCTTCCCATCTTTCAGACCTGGGTGATTAACATCACCCGAGCAG 143543
QY 1067 GGCATGGGTGATCCAGCTTTTCCACACAAATGCGCAGGAGATCCCTTAATGTCTCCAGGCT 1126
Db 143544 GGCATGGGTGATCCAGCTTTTCCACACAAATGCGCAGGAGATCCCTTAATGTCTCCAGGCT 143603
QY 1127 TGGGCGTGGCTTTGACCTTATGAAAAGTTTCTGACTTTCAGAAAGCTGTGGGAGTGC 1186
Db 143604 TGGGCGTGGCTTTGACCTTATGAAAAGTTTCTGACTTTCAGAAAGCTGTGGGAGTGC 143663
QY 1187 ATGAATTTGCTGGGCTACCGCCAGCTGATCTGTAACAGAAACAGAAAACCTGTTGCTG 1246
Db 143664 ATGAATTTGCTGGGCTACCGCCAGCTGATCTGTAACAGAAACAGAAAACCTGTTGCTG 143723
QY 1247 GATCTTCTGCTACCTGAGATCTGCTCCAGCAATTCACATTAAGGGTTTGAAGGCTTT 1306
Db 143724 GATCTTCTGCTACCTGAGATCTGCTCCAGCAATTCACATTAAGGGTTTGAAGGCTTT 143783
QY 1307 GCTGCCACTGTGTGACCTTCAGATCACTTCTCTGAATGCTTGAAGCTTGGCTTCAT 1366
Db 143784 GCTGCCACTGTGTGACCTTCAGATCACTTCTCTGAATGCTTGAAGCTTGGCTTCAT 143843
QY 1367 CTCTGAGCCCTTAACTATATCTGTGTGGGTATCACACTGAGTGTGATGTCACACAGT 1426
|||||

Db 143844 CTCTGAGCCCTTAACTATATCTGTGTGGGTATCACACTGAGTGTGATGTCACACAGT 143903
QY 1427 GCTCAAGCAGAGAGCTTTGTGTCATGCTGTGCTAGAAAACAGAGTGGGAACCTT 1486
Db 143904 GCTCAAGCAGAGAGCTTTGTGTCATGCTGTGCTAGAAAACAGAGTGGGAACCTT 143963
QY 1487 ATGTGACAGACACATCTCCACAGTGAAGAGGATGCTTCTTCTTCTTCTGATCTT 1546
Db 143964 ATGTGACAGACACATCTCCACAGTGAAGAGGATGCTTCTTCTTCTTCTGATCTT 144023
QY 1547 CCTGTCTGGGACACTTCAGAGACTTGTGTGGCTGTGAGGCTTATTAAGCACAGACAGTA 1606
Db 144024 CCTGTCTGGGACACTTCAGAGACTTGTGTGGCTGTGAGGCTTATTAAGCACAGACAGTA 144083
QY 1607 TCACTGGAATTTGATCAATTAACCTCCCTGTCCACATCTTGGCCAAATGGGAATGATCTT 1666
Db 144084 TCACTGGAATTTGATCAATTAACCTCCCTGTCCACATCTTGGCCAAATGGGAATGATCTT 144143
QY 1667 TCACCAAGAGAGCTCCACAGATTTTCCACAGAGATGCGAAATTCGAGCCCTTGGAGTTC 1726
Db 144144 TCACCAAGAGAGCTCCACAGATTTTCCACAGAGATGCGAAATTCGAGCCCTTGGAGTTC 144203
QY 1727 CAATGGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1786
Db 144204 CAGT-GGATTCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 144262
QY 1787 T-ACAGTATCGGTAATCAGAAATATGAACAAATATCTG-ACAAAGAGAGAGAGAGAG 1844
Db 144263 TCACAGTATCGGTAATCAGAAATATGAACAAATATCTG-ACAAAGAGAGAGAGAGAG 144322
QY 1845 AAGTTCAAGAGTCCCTGGGCTTGTGATGATATCA-TTTCCTTTCATTTTCCATTA 1903
Db 144323 AAGTTCAAGAGTCCCTGGGCTTGTGATGATATCA-TTTCCTTTCATTTTCCATTA 144382
QY 1904 CATGAAACTTTGACCTGTGAACCTTGCATCTGTTAATCAATTAATTTCCAAATAGG 1963
Db 144383 CATGAAACTTTGACCTGTGAACCTTGCATCTGTTAATCAATTAATTTCCAAATAGG 144441
QY 1964 TTTCTTTTAAATTCCTTTTATGCTTCTTATTTTAAAGCAATTAATGTCATTTT 2023
Db 144442 TTTCTTTTAAATTCCTTTTATGCTTCTTATTTTAAAGCAATTAATGTCATTTT 144501
QY 2024 TGGGATCTTAAAAAAA 2042
Db 144502 TGGGATCTTAAAGTAAGAA 144520

RESULT 7
AC138848/c 206943 bp DNA linear HTG 21-JAN-2003
LOCUS
DEFINITION
Homo sapiens chromosome 5 clone RP11-1301B21, WORKING DRAFT
SEQUENCE, 7 unordered pieces.
AC138848
AC138848.1 GI:27805260
VERSION
HTG: HTGS-PHASE1; HTGS-DRAFT; HTGS-ACTIVEFIN.
KEYWORDS
Homo sapiens (human)
SOURCE
Homo sapiens
ORGANISM
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE
1 (bases 1 to 206943)
DOE Joint Genome Institute.
Unpublished
Sequencing of Human Chromosome 5
2 (bases 1 to 206943)
DOE Joint Genome Institute.
Direct Submission
Submitted (21-JAN-2003) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
```

Project Information
Center Project Name: 2719263
Center clone name: RPII-11.1301B21

Summary Statistics
Consensus quality: 202039 bases at least Q40
Consensus quality: 202804 bases at least Q30
Consensus quality: 203337 bases at least Q20
Estimated insert size: 175000; agarose-fp estimation
Estimated insert size: 206343; sum-of-coverage estimation
Quality coverage: 15.28 in Q20 bases; agarose-fp estimation
Quality coverage: 12.96 in Q20 bases; sum-of-coverage estimation.
NOTE: This is a 'working draft' sequence. It currently
consists of 7 contigs. The true order of the pieces
is not known and their order in this sequence record is
arbitrary. Gaps between the contigs are represented as
runs of N, but the exact sizes of the gaps are unknown.
This record will be updated with the finished sequence
as soon as it is available and the accession number will
be preserved.

1 1219: contig of 1219 bp in length
* 1220 1319: gap of unknown length
* 1320 2493: contig of 1174 bp in length
* 2494 2593: gap of unknown length
* 2594 3666: contig of 1073 bp in length
* 3667 3766: gap of unknown length
* 3767 5167: contig of 1401 bp in length
* 5168 5267: gap of unknown length
* 5268 12818: contig of 7551 bp in length
* 12819 12918: gap of unknown length
* 12919 36169: contig of 23251 bp in length
* 36170 36269: gap of unknown length
* 36270 206943: contig of 170674 bp in length.

Location/Qualifiers
1. 206943
/organism="Homo sapiens"
/mol_type="genomic DNA"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RPII-1301B21"
/clone_lib="RPII human BMC library 11"

BASE COUNT 56312 a 44122 c 46163 g 59225 t 1121 others

ORIGIN

Query Match 90.9%; Score 1856.6; DB 2; Length 206943;
Best Local Similarity 99.0%; Pred. No. 0;
Matches 1920; Conservative 0; Mismatches 14; Indels 5; Gaps 5;

107 AAGGCTTCCACTTCCAGCAATGCTACTGCTTAAATAAATGAAGCTCGTGTTCG 166
125368 AAGGCTTCCACTTCCAGCAATGCTACTGCTTAAATAAATGAAGCTCGTGTTCG 125309

167 GTTTCAGATGAGGCAATCTTGGCTATTTCCACATGTACAGCCACAACATCAGCTCC 226
125308 GTTTCAGATGAGGCAATCTTGGCTATTTCCACATGTACAGCCACAACATCAGCTCC 125249

227 CTTGTATTAAGGCAAGCCGAGGCAATGCTGCTGCTTCTGCTGCGCTCT 286
125248 CTTGTATTAAGGCAAGCCGAGGCAATGCTGCTGCTTCTGCTGCGCTCT 125189

287 GGTCTCTCTTTTGGGGAGGCTTTTGGGAGCAACAGATGTTTTCACCGATGAG 346
125188 GGTCTCTCTTTTGGGGAGGCTTTTGGGAGCAACAGATGTTTTCACCGATGAG 125129

347 CCCGCTGAGCAGTGTGATGACCTTCAAGCAGAGCAGCCGCTGATGCTGCATGCT 406
125128 CCCGCTGAGCAGTGTGATGACCTTCAAGCAGAGCAGCCGCTGATGCTGCATGCT 125069

407 GTGCGGATCTGATAGAGGCGCTCTTGTGCGAGATGAGCGCTTTGATGCTATG 466
125068 GTGCGGATCTGATAGAGGCGCTCTTGTGCGAGATGAGCGCTTTGATGCTATG 125009

467 GAACCTGTCCTCCGAGAGACAGTCCAGCTCTTCACTGAGGAGAACAGCCGCTGTGT 526

125008 GAACCTGTCCTCCGAGAGACAGTCCAGCTCTTCACTGAGGAGAACAGCCGCTGTGT 124949
527 TCTGACAGTCCCTGTGACATCATCCACAAGATGAATATCCCCGGGCTCATGTCAG 586
124948 TCTGACAGTCCCTGTGACATCATCCACAAGATGAATATCCCCGGGCTCATGTCAG 124889

587 CTCCTGTGACATCATCCAGCTTTGAGGTGTGAGAGAGGCTCCGCTCTCAACAGCC 646
124888 CTCCTGTGACATCATCCAGCTTTGAGGTGTGAGAGAGGCTCCGCTCTCAACAGCC 124829

647 GTGTGCTCAAGAGAGTGGCTTTCTTCAACCTGACATCCCTCTACCCGCTGTAAGAG 706
124828 GTGTGCTCAAGAGAGTGGCTTTCTTCAACCTGACATCCCTCTACCCGCTGTAAGAG 124769

707 CCGTCCCTCAACCTGATATCGTACCTGTCGTCGAGGAGCCCGGGCCGTGTTCCGTC 766
124768 CCGTCCCTCAACCTGATATCGTACCTGTCGTCGAGGAGCCCGGGCCGTGTTCCGTC 124709

767 CGAGAAAGCAAAAGGAGATCTCATGATGACAGTGCATTTGTATGGGAGCATGAG 826
124708 CGAGAAAGCAAAAGGAGATCTCATGATGACAGTGCATTTGTATGGGAGCATGAG 124649

827 CAAAACCTAAGAGAGAGACCAACCTCTATGATGATGATGATGATGATGATGATGAT 886
124648 CAAAACCTAAGAGAGAGAGACCAACCTCTATGATGATGATGATGATGATGATGATGAT 124589

887 CTGAGATCTACAAAGCAATCCAGTCTTGGCCAAAGGCTCGAGAGAGCTACCTGCTT 946
124588 CTGAGATCTACAAAGCAATCCAGTCTTGGCCAAAGGCTCGAGAGAGCTACCTGCTT 124529

947 GTGCGCTATGAGAGCTGGCTGAGAGCCCTGTGAGCCCAAGCTCCGAATGATGATTC 1006
124528 GTGCGCTATGAGAGCTGGCTGAGAGCCCTGTGAGCCCAAGCTCCGAATGATGATTC 124469

1007 GTGAGATGAATATTTGGCCCATCTTCAAGACCTGGGTGATTAATCAACCCAGAGCAAG 1066
124468 GTGAGATGAATATTTGGCCCATCTTCAAGACCTGGGTGATTAATCAACCCAGAGCAAG 124409

1067 GGCATGATGAGACAGCTTTCCACACAAATGCGAGGATGCCCTTAATGCTCCAGGCT 1126
124408 GGCATGATGAGACAGCTTTCCACACAAATGCGAGGATGCCCTTAATGCTCCAGGCT 124349

1127 TGGCGCTGCTTTGGCCCATGAAAGGTTTCTGCACTTCAAGAAACCTGTCGCTGCT 1186
124348 TGGCGCTGCTTTGGCCCATGAAAGGTTTCTGCACTTCAAGAAACCTGTCGCTGCT 124289

1187 ATGAATTTCTGAGGCTAACCGCAAGTCAATGATGATGATGATGATGATGATGATGAT 1246
124288 ATGAATTTCTGAGGCTAACCGCAAGTCAATGATGATGATGATGATGATGATGATGAT 124229

1247 GATCTTCTGCTACCTGAGCTGCTCCAGCAATTCACATTAAGAGGCTTGAAGAGCTTT 1306
124288 GATCTTCTGCTACCTGAGCTGCTCCAGCAATTCACATTAAGAGGCTTGAAGAGCTTT 124169

1307 GCTGCGACCTGGTGTAGCTCAGTACCTTCTCTAATGCTTCTGAGCTTGGCTACAT 1366
124168 GCTGCGACCTGGTGTAGCTCAGTACCTTCTCTAATGCTTCTGAGCTTGGCTACAT 124109

1367 CTTGTGAGCTTAACTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1426
124108 CTTGTGAGCTTAACTATGATGATGATGATGATGATGATGATGATGATGATGATGAT 124049

1427 GCTCAAGCAGAGGATTTTGTGCTCAGCTTGTGCTAGAAACAGACTGGGGAACCTT 1486
124048 GCTCAAGCAGAGGATTTTGTGCTCAGCTTGTGCTAGAAACAGACTGGGGAACCTT 123989

1487 ATGTGAGCAGCAATCCACAGTGAAGAGGATTTGCTCTTCTTCTTGTGATCTT 1546
123988 ATGTGAGCAGCAATCCACAGTGAAGAGGATTTGCTCTTCTTCTTGTGATCTT 123929

1547 CCTGTGAGGAGACTTCAAGAGACTTGTGGCTTGAAGGCTTATTAAGACAGACAGTA 1606
123929 CCTGTGAGGAGACTTCAAGAGACTTGTGGCTTGAAGGCTTATTAAGACAGACAGTA 1606

THIS PAGE BLANK (USPTO)

Db 1276 CATACTAACGCCAGAACGCCCTCAACGTCCTCAGCGGTGGTCCCTTACTTAC 1335
OY 1148 GAAAGGTTCTGACTTCAGAAAGCCCTGGCCATGCCATTAATTTCTGGGCTACCGC 1207
Db 1336 GAAAGGTTCTGACTTCAGAAAGCCCTGGCCATGCCATTAATTTCTGGGCTACCGC 1395
OY 1208 CACGTCAGATCTGAACAAGAACAGAAAGCTGTGGATCTTCTGTCTACTGTGACT 1267
Db 1396 CAGGTCAGATCTGAACAAGAACAGAAAGCTGTGGATCTTCTGTCTACTGTGACT 1455
OY 1268 GTCCCTGAGCAATCCACTAAGAGGTTGAGAAGCTTGTCTG--CCACTGGTGTACG 1325
Db 1456 ATCTTGGGGCAGGCTTCCGAGAGGTTAAGAGGTCGTCTGCACCCCTGGTCCAGC 1515
OY 1326 CTCAGTCACCTTCTGTGAATGCTTGTGAGCCTTGCCCTACATCTGAGCCCTTAATACAT 1385
Db 1516 CTTTGTCTAC--CATTAACGCAAGAGCCCTTAAGGTAGAACCAACTGAGTGCCCTT 1572
OY 1386 GTCGTGGGTATCACACTGAGTGTGATTTGTCCACAGTGTCAAGCAGAGGACTTT 1445
Db 1573 TCTCCTCAGCCCAAGCAGAGAGGCTTGTGTCTATCTCAGGTCCTACCTACACTGA 1632
OY 1446 TGTGTCCATGCTTGTGTCTAGAAAACAGACTGGGAACTTATGTGACGACACATCCA 1505
Db 1633 GCCTAAAAAGCCAAAGAAACATATCTTCTGTGAATAATTAAGAACCTTAAGCAG 1692
OY 1506 CCAGTGAACAGGGTA---TTGCTCTCTCTTCTTCTGATCTCTCTGCTGGGAGACT 1562
Db 1693 CCCCTTGACCTGTCAAGCAAGACTTCTTGAATTTGGCCTTCTTACTGTGCAATACC 1752
OY 1563 TCAGAGACTTGTGGCCCTGAGGCTTATTAAAGACAGACAGTATCACTGGAATTTGATCC 1622
Db 1753 TTGGAGACTAGGCTGTGAGGCTTACTGACACAGCAAGCAATCCATGAGTGTCT 1812
OY 1623 ATTAACCTCCCTGTCCATCTTGTGCCAATGGGAATGATCTTCAACCAAGACTCAC 1682
Db 1813 GTAAACCTCCCTGTCCATCTTGTCCGAAAGAAAGGGAGAAATAAAGCCCAAGGAAATG 1872
OY 1683 CAGCATTTTCCACAGAGATGCGA-----ATTCTGAGCCCTTGGAGTTCCCA 1729
Db 1873 CGGCTCTTACCAAGAGCTCCCAAGAACTTCCACAGGACATAGACCTCAAGTTTCCA 1932
OY 1730 TGGATTCAGGAAGAAAGTGGAAACAAGTTGATGCTTACTTATGAGCTTGACCATAC 1789
Db 1933 GCGGATTTGAGAGAGAGGAGCGGGGTGGGTATCT--GTGAATCTGTGCCATGT 1991
OY 1790 AGCTATGCTAATGAAATATGAACAATAATCTGTGCAAAAGCAAGCTTAATTT 1849
Db 1992 CACTATCCGTATACAGAAACAGAAAGCAAA---CAACACAAAGGGGCAAGACTCAAGTT 2048
OY 1850 CACAAGTGCTGGGCTTGAT 1870
Db 2049 CAAGGCTGTCTGGTATGAT 2069

Search completed: August 14, 2003, 15:34:38
Job time : 7298.74 secs

FEATURES FT /organism='Homo sapiens (human)'.
 source 1. .877
 Location/Qualifiers
 /organism='Homo sapiens'
 /mol_type='genomic DNA'
 /db_xref='taxon:9606'

BASE COUNT 183 a 266 c 221 g 203 t 4 others
 ORIGIN

Query Match 37.4% Score 763.6; DB 6; Length 877;
 Best Local Similarity 97.7%; Pred. No. 5.3e-182;
 Matches 795; Conservative 0; Mismatches 16; Indels 3; Gaps 2;

94 AAGCCAAAGCAGAGTCTTCCACTTCAGCAATGCTACTGCTTAAATAATGAACT 153
 |||||
 35 AAGCCGCTTGAAGGCTTCCACTTCAGCAATGCTACTGCTTAAATAATGAACT 94
 |||||
 154 CCTGCTGTTTGTGTTTCCAGATGCGCATCTTGCTCTATTCTTCCACATGTACGCCA 213
 |||||
 95 CCGCTGTTTGTGTTTCCAGATGCGCATCTTGCTCTATTCTTCCACATGTACGCCA 154
 |||||
 214 CAACATGAGTCCCTGCTATGAGGACAGCCGAGGCAATGCACTGCTGCTCTGTC 273
 |||||
 155 CAACATGAGTCCCTGCTATGAGGACAGCCGAGGCAATGCACTGCTGCTGTC 214
 |||||
 274 TTCTGCGGCTCTGCTCTCTTTTGTGCGGAGCTTTTGGGAGCACCAGATGTTT 333
 |||||
 215 TTCTGCGGCTCTGCTCTCTTTTGTGCGGAGCTTTTGGGAGCACCAGATGTTT 274
 |||||
 334 CTACCTGATGAGCCCGCTGCGACGTGTGATGACCTTCAAGCAGACCCGCTGAT 393
 |||||
 275 CTACCTGATGAGCCCGCTGCGACGTGTGATGACCTTCAAGCAGACCCGCTGAT 334
 |||||
 394 GCTGCACTGAGCTGTGCGGATGATGATGAGGCGGCTCTTGTGCGCATGAGGCTT 453
 |||||
 335 GCTGCACTGAGCTGTGCGGATGATGATGAGGCGGCTCTTGTGCGCATGAGGCTT 394
 |||||
 454 TGATGCTACATGAGAACCTGTGTCGCCGAGACAGTCCAGCCCTTTCACTGAGGAGAACAG 513
 |||||
 395 TGATGCTACATGAGAACCTGTGTCGCCGAGACAGTCCAGCCCTTTCACTGAGGAGAACAG 454
 |||||
 514 CCGGGCCCTGTGTTCTGACCTGCTGTGACATATCCCAAGATGAATCATCCCGG 573
 |||||
 455 CCGGGCCCTGTGTTCTGACCTGCTGTGACATATCCCAAGATGAATCATCCCGG 514
 |||||
 574 GGTCTACTGAGGCTCTGTCAGATCAACAGCCCTTGTGAGTGGGAGAGGCGCTGCG 633
 |||||
 515 GGTCTACTGAGGCTCTGTCAGATCAACAGCCCTTGTGAGTGGGAGAGGCGCTGCG 574
 |||||
 634 CTCTACAGCAGTGTGTCAGAGAGGCTGCTTTCAACCTGACAGTCCCTCTACCC 693
 |||||
 575 CTCTACAGCAGTGTGTCAGAGAGGCTGCTTTCAACCTGACAGTCCCTCTACCC 634
 |||||
 694 GCTGCTGAAGACCCCTCCCAACCTGATATGCTGACCTGTCGCCGAGCCCGGCG 753
 |||||
 635 GCTGCTGAAGACCCCTCCCAACCTGATATGCTGACCTGTCGCCGAGCCCGGCG 694
 |||||
 754 CGTGTTCGTTCCGAGAGCAGCAAGAGGAGATCTCATGTGACATTCGATTTGNAT 754
 |||||
 695 CGTGTTCGTTCCGAGAGCAGCAAGAGGAGATCTCATGTGACATTCGATTTGNAT 754
 |||||
 814 GGGGAGCAGTGAAGAACTCAAGAGAGGAGACCACTTATGTGATGACAGTCA - 872
 |||||
 755 GGGGAGCAGTGAAGAACTCAAGAGAGGAGACCACTTATGTGATGACAGTCA - 872
 |||||
 873 -TTCGCAAGACGAGC-TGAGATCTTACAGACC 904
 |||||
 815 TCTGCCAAGACGAGCTTGAGAACTTACAGAAC 848
 |||||

DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 ACCESSION BD126950
 VERSION BD126950.1 GI:23221895
 KEYWORDS JP 2002017375-A/2381.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homidae; Homo.
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
 Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
 Koga,H.
 Primer for synthesizing full-length cDNA and use thereof
 Patent: JP 2002017375-A 2381 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/2381
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,AI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUJI OTSUKI,HISASHI KOGA
 PC C12N15/09,C07K14/47,C07K16/18,C12N1/15,C12N1/19,C12N1/21,C12N5/PC
 10.
 PC C12P1/02,C12Q1/68//C12P1/08,G06F17/30,C12N15/00,C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FI Key

FEATURES FT /organism='Homo sapiens (human)'.
 source 1. .877
 Location/Qualifiers
 /organism='Homo sapiens'
 /mol_type='genomic DNA'
 /db_xref='taxon:9606'

BASE COUNT 183 a 266 c 221 g 203 t 4 others
 ORIGIN

Query Match 37.4% Score 763.6; DB 6; Length 877;
 Best Local Similarity 97.7%; Pred. No. 5.3e-182;
 Matches 795; Conservative 0; Mismatches 16; Indels 3; Gaps 2;

94 AAGCCAAAGCAGAGTCTTCCACTTCAGCAATGCTACTGCTTAAATAATGAACT 153
 |||||
 35 AAGCCGCTTGAAGGCTTCCACTTCAGCAATGCTACTGCTTAAATAATGAACT 94
 |||||
 154 CCTGCTGTTTGTGTTTCCAGATGCGCATCTTGCTCTATTCTTCCACATGTACGCCA 213
 |||||
 95 CCGCTGTTTGTGTTTCCAGATGCGCATCTTGCTCTATTCTTCCACATGTACGCCA 154
 |||||
 214 CAACATGAGTCCCTGCTATGAGGACAGCCGAGGCAATGCACTGCTGCTCTGTC 273
 |||||
 155 CAACATGAGTCCCTGCTATGAGGACAGCCGAGGCAATGCACTGCTGCTGTC 214
 |||||
 274 TTCTGCGGCTCTGCTCTCTTTTGTGCGGAGCTTTTGGGAGCACCAGATGTTT 333
 |||||
 215 TTCTGCGGCTCTGCTCTCTTTTGTGCGGAGCTTTTGGGAGCACCAGATGTTT 274
 |||||
 334 CTACCTGATGAGCCCGCTGCGACGTGTGATGACCTTCAAGCAGACCCGCTGAT 393
 |||||
 275 CTACCTGATGAGCCCGCTGCGACGTGTGATGACCTTCAAGCAGACCCGCTGAT 334
 |||||
 394 GCTGCACTGAGCTGTGCGGATGATGATGAGGCGGCTCTTGTGCGCATGAGGCTT 453
 |||||
 335 GCTGCACTGAGCTGTGCGGATGATGATGAGGCGGCTCTTGTGCGCATGAGGCTT 394
 |||||
 454 TGATGCTACATGAGAACCTGTGTCGCCGAGACAGTCCAGCCCTTTCACTGAGGAGAACAG 513
 |||||
 395 TGATGCTACATGAGAACCTGTGTCGCCGAGACAGTCCAGCCCTTTCACTGAGGAGAACAG 454
 |||||
 514 CCGGGCCCTGTGTTCTGACCTGCTGTGACATATCCCAAGATGAATCATCCCGG 573
 |||||

/organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 93..1253
 /note="LSTR"
 /codon_start=1
 /product="L-selectin ligand sulfotransferase"
 /protein_id="AAK48417.1"
 /db_xref="GI:13897504"
 /translation="MLPKMKILFLIVSOMAILAFPMVSHNISLSMKRQPRMH
 VLIVSMRSGSFVQLEQGHDPVFLMEPANHVMITKOSTAMHLHVAVDLITAVE
 LCDMSVDAIMEPEPRROSLFOWENSRALCSAPACDIIPODEIIPRAHRLCSQOP
 FEVEKACRSYSHVLAKEVREFNLSLPLKDPSLNHLIYLVDPRAVRSRRTK
 GDLMDISRIVMGHOHLKLEDOPIYVQVICOSELEIKYKIOSLPKALOERYLIVRY
 EDIAPAPVOTSRMEFVLEFLPHLOIWMVHNITRGKGGDIAFPTNADLANVSQAW
 RMSLPYEVSRLOKACGDAMNLIQYRHRSEDEQRNLLDLISTVTEYEQIH"

BASE COUNT 286 a 393 c 336 g 318 t
 ORIGIN

Query Match 61.8%; Score 1262; DB 9; Length 1333;
 Best Local Similarity 100.0%; Pred. No. 3.6e-308;
 Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

107 AAGGCTTCCACTTATGACCAATGCTACTGCTAAAAAATAGCTCCTGCTTTCTG 166
 72 AAGGCTTCCACTTATGACCAATGCTACTGCTAAAAAATAGCTCCTGCTTTCTG 131
 167 GTTTCACAGATGGCATCTGGCTATTTCTTCACATGTACAGCACAACATACAGCTCC 226
 132 GTTTCACAGATGGCATCTGGCTATTTCTTCACATGTACAGCACAACATACAGCTCC 191
 227 CTGTCTATGAGGACAGCCGAGCGCATGCTGCTGTTCTGTCTTCTGCGGCTCT 286
 192 CTGTCTATGAGGACAGCCGAGCGCATGCTGCTGTTCTGTCTTCTGCGGCTCT 251
 287 GGCTCTCTTGTGGGGAGCTTTTGGGACAGCCAGATGTTTCTACTGATGAGG 346
 252 GGCTCTCTTGTGGGGAGCTTTTGGGACAGCCAGATGTTTCTACTGATGAGG 311
 347 CCCGCTGACAGCTGTGATGACCTTCAAGACAGACAGCCGCTGATGCTGACATAGCT 406
 312 CCCGCTGACAGCTGTGATGACCTTCAAGACAGACAGCCGCTGATGCTGACATAGCT 371
 407 GTGCGGATCTATACGGGCGCTCTTCTTGTGCGACATGAGCGCTTGTGATGCTTACATG 466
 372 GTGCGGATCTATACGGGCGCTCTTCTTGTGCGACATGAGCGCTTGTGATGCTTACATG 431
 467 GAACCTGGTCCCGGAGACATGACGCTTTCAGTGGGAGAAACAGCCGGGCGCTGTGT 526
 432 GAACCTGGTCCCGGAGACATGACGCTTTCAGTGGGAGAAACAGCCGGGCGCTGTGT 491
 527 TCTGACACCTGCTGTGACATATCCACAAAGATGAATATATCCCGGGGCTCACTGACAG 586
 492 TCTGACACCTGCTGTGACATATCCACAAAGATGAATATATCCCGGGGCTCACTGACAG 551
 587 CTCCTGTGACATCAACAGCCCTTTGAGGTGTGAGAGAGCGCTCCGCTCTCAACAGCCAC 646
 552 CTCCTGTGACATCAACAGCCCTTTGAGGTGTGAGAGAGCGCTCCGCTCTCAACAGCCAC 611
 647 GTGGGTCTCAAGGAGGTGGCTTCTTCAACGTCGACGCTCCCTACCCGCTGCTGAAGAC 706
 612 GTGGGTCTCAAGGAGGTGGCTTCTTCAACGTCGACGCTCCCTACCCGCTGCTGAAGAC 671
 707 CCCTCCCTCAACCTGATATGTCGACATGCTCCGGAGACCCCGGGGCTGTTCCGCTTC 766
 672 CCCTCCCTCAACCTGATATGTCGACATGCTCCGGAGACCCCGGGGCTGTTCCGCTTC 731
 767 CGAGAACGACAAAGGAGATCTATGATGACAGTGTGATGATGGGAGAGATGAG 826
 732 CGAGAACGACAAAGGAGATCTATGATGACAGTGTGATGATGGGAGAGATGAG 791
 827 CAAAACCTCAAGAGAGAGACCAACCTTACTATGTGATCAGAGTATCTGCGCAAGCCAG 886

Db 792 CAAAACCTCAAGAGAGAGACCAACCTTACTATGTGATGACAGTATCTGCGCAAGCCAG 851
 Oy 887 CTGGAGATCTAACAAGACATCCAGTCTTGGCCCAAGCCCTCAGAGAACGTCACCTGCT 946
 Db 852 CTGGAGATCTAACAAGACATCCAGTCTTGGCCCAAGCCCTCAGAGAACGTCACCTGCT 911
 Oy 947 GTGCGCTATGAGACCTGCTGACAGCCCTGTGGCCCAAGCTTCCGAAATGATGAATTC 1006
 Db 912 GTGCGCTATGAGACCTGCTGACAGCCCTGTGGCCCAAGCTTCCGAAATGATGAATTC 971
 Oy 1007 GTGGAGTTGGAATTTCTTGGCCCATCTTCAAGACCTGGGTGATTAATCAACCCGAGCAAG 1066
 Db 972 GTGGAGTTGGAATTTCTTGGCCCATCTTCAAGACCTGGGTGATTAATCAACCCGAGCAAG 1031
 Oy 1067 GGCATGGGAGACAGCTTCCACCAAAATGGCAGAGGATGCCCTTAATGTCGCCAGCT 1126
 Db 1032 GGCATGGGAGACAGCTTCCACCAAAATGGCAGAGGATGCCCTTAATGTCGCCAGCT 1091
 Oy 1127 TGGCGCTGCTCTTGGCCCATATGAAAAGTTTCTGACACTTCAGAAAAGCTGTGGCAGATGC 1186
 Db 1092 TGGCGCTGCTCTTGGCCCATATGAAAAGTTTCTGACACTTCAGAAAAGCTGTGGCAGATGC 1151
 Oy 1187 ATGAATTTGCTGGGCTACCGCCACGTCGACATCTGAACAAGAACAGAAACCTGTTGCTG 1246
 Db 1152 ATGAATTTGCTGGGCTACCGCCACGTCGACATCTGAACAAGAACAGAAACCTGTTGCTG 1211
 Oy 1247 GATCTTCTGCTACCTGACGACTGCTCCCTGAGCAATCCATAGAGAGGTTGAGAAAGCTTT 1306
 Db 1212 GATCTTCTGCTACCTGACGACTGCTCCCTGAGCAATCCATAGAGAGGTTGAGAAAGCTTT 1271
 Oy 1307 GCTGCACCTGCTGTCAGCCTGACGCTTCTGTAATGCTTGTGAGGCTTGCTTACAT 1366
 Db 1272 GCTGCACCTGCTGTCAGCCTGACGCTTCTGTAATGCTTGTGAGGCTTGCTTACAT 1331
 Oy 1367 CT 1368
 Db 1332 CT 1333

RESULT 13
 BD124832
 LOCUS 877 bp DNA linear PART 18-SEP-2002
 DEFINITION Primer for synthesizing full-length cDNA and use thereof.
 VERSION BD124832.1 GI:23219777
 KEYWORDS JP 2002017375-A/263.
 SOURCE Homo sapiens (human)
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 877)
 Ota,T., Nishikawa,T., Isogai,T., Hayashi,K., Ishii,S., Kawai,Y.,
 Wakamatsu,A., Sugiyama,T., Nagai,K., Kojima,S., Otsuki,T. and
 Koga,H.
 Patent: JP 2002017375-A 263 22-JAN-2002;
 HELIX RESEARCH INSTITUTE
 OS Homo sapiens (human)
 PN JP 2002017375-A/263
 PD 22-JAN-2002
 PF 07-JUL-2000 JP 2000253172
 PI TOSHIO OTA,TETSUO NISHIKAWA,TAKAO ISOGAI,KOJI HAYASHI,SHIZUKO
 PI ISHII,
 PI YURI KAWAI,MI WAKAMATSU,TOMOYASU SUGIYAMA,KEIICHI NAGAI, PI
 SHINICHI KOJIMA,
 PI TETSUO OTSUKI,HISASHI KOA

TITLE
 JOURNAL
 COMMENT
 PC C12N15/09, C07R14/47, C07R16/18, C12N1/15, C12N1/19, C12N1/21, C12N5/ PC
 10, C12P21/02, C12Q1/68, C12P21/08, G06F17/30, C12N15/00, C12N5/00 CC
 Primer for synthesizing full-length cDNA and use thereof FH Key
 Location/Qualifiers
 FT source I. 877

Db 268 CAACATAGCTCCCTGCTATGAGGACAGCCGACGATGACCTGCTGCTGTC 327
 QY 274 TTCTGCGGCTGCGCTCTTCTTGTGGGACAGCTTTGGGAGACCCAGATGTTT 333
 Db 328 TTCTGCGGCTGCTGCTCTTCTTGTGGGACAGCTTTGGGAGACCCAGATGTTT 387
 QY 334 CTACCTGATGAGAGCCGCTGACAGCTGTGGATGACCTTCAAGAGACAGCCGCTGAT 393
 Db 388 CTACCTGATGAGAGCCGCTGACAGCTGTGGATGACCTTCAAGAGAGACCCGCTGAT 447
 QY 394 GCTGCAATGCTGTGGGATGATGATGAGGCGCTTCTTGTGCGACATGACCTTCT 453
 Db 448 GCTGCAATGCTGTGGGATGATGATGAGGCGCTTCTTGTGCGACATGACCTTCT 507
 QY 454 TGAATGCTGATGAGAGCTGCTGCGGAGACAGCTGACCTTGTAGTGGAGAACAG 513
 Db 508 TGAATGCTGATGAGAGCTGCTGCGGAGACAGCTGACCTTGTAGTGGAGAACAG 567
 QY 514 CCGGGCCCTGTGTTCTGACCTGCTGACATCATCCCAAGATGAATCATCCCG 573
 Db 568 CCGGGCCCTGTGTTCTGACCTGCTGACATCATCCCAAGATGAATCATCCCG 627
 QY 574 GGTCTACTGAGGCTCTGTGCACTCAACAGCCCTTTGAGTGTGGAGAGCCCTGCG 633
 Db 628 GGTCTACTGAGGCTCTGTGCACTCAACAGCCCTTTGAGTGTGGAGAGCCCTGCG 687
 QY 634 CTCTACAGCAGCGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGAGTCCCTTACC 693
 Db 688 CTCTACAGCAGCGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGAGTCCCTTACC 747
 QY 694 GCTGCTGAAAGAGCCCTCTCAACCTGATCTGTCACCTGCTGCGGAGCCCGGCG 753
 Db 748 GCTGCTGAAAGAGCCCTCTCAACCTGATCTGTCACCTGCTGCGGAGCCCGGCG 807
 QY 754 CGTGTCCGTTCCCGAGAGACGACAAAGGATCTGATGTTGACAGTGGCTGAT 813
 Db 808 CGTGTCCGTTCCCGAGAGACGACAAAGGATCTGATGTTGACAGTGGCTGAT 867
 QY 814 GGGGACATGAG 873
 Db 868 GGGGACATGAG 927
 QY 874 CTGCCAAGCAGCTGAGATCTACAGAGACATCTGCTGCGGAGCCCTGACAGA 933
 Db 928 CTGCCAAGCAGCTGAGATCTACAGAGACATCTGCTGCGGAGCCCTGACAGA 987
 QY 934 ACCTACCTGCTGTGCTATGAGAGAGCTGCTGAGCCCTGCGGAGAGCTTCCG 993
 Db 988 ACCTACCTGCTGTGCTATGAGAGAGCTGCTGAGCCCTGCGGAGAGCTTCCG 1047
 QY 994 AATGTATGATGTGGGATGGAATCTTGGCCCATCTTCAAGCCCTGGGAGCTTAA 1053
 Db 1048 AATGTATGATGTGGGATGGAATCTTGGCCCATCTTCAAGCCCTGGGAGCTTAA 1107
 QY 1054 CACCCGAG 1113
 Db 1108 CACCCGAG 1167
 QY 1114 TGTCTCCAGAGCTTGGGCTGTGCTTGGCCATGAAAGGTTTCTGACTTCAAGAG 1173
 Db 1168 TGTCTCCAGAGCTTGGGCTGTGCTTGGCCATGAAAGGTTTCTGACTTCAAGAG 1227
 QY 1174 CTGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1233
 Db 1228 CTGTGGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1287
 QY 1234 AATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1293
 Db 1288 AATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1347
 QY 1294 TTGAGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1353

Db 1348 TTGAGAGGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1407
 QY 1354 GCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1413
 Db 1408 GCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1467
 QY 1414 TGTGTCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1473
 Db 1468 TGTGTCACAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1527
 QY 1474 ACTGGGAGCTTATGAG 1533
 Db 1528 ACTGGGAGCTTATGAG 1587
 QY 1534 TTTCTGTGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1593
 Db 1588 TTTCTGTGATCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1647
 QY 1594 GCACGACACAGTATCACTGATGATGATGATGATGATGATGATGATGATGATGATG 1653
 Db 1648 GCACGACACAGTATCACTGATGATGATGATGATGATGATGATGATGATGATGATG 1707
 QY 1654 GGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1713
 Db 1708 GGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1767
 QY 1714 CCTTGGAGTCCCATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1773
 Db 1768 CCTTGGAGTCCCATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATG 1826
 QY 1774 ATGAGCTTGACAT -ACAGTATGCTGATGATGATGATGATGATGATGATGATGATG 1831
 Db 1827 ATGAGCTTGACATGACATGATGATGATGATGATGATGATGATGATGATGATGATG 1886
 QY 1832 AGACCAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1890
 Db 1887 AGACCAAGCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1946
 QY 1891 CATTTCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1950
 Db 1947 CATTTCCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 2005
 QY 1951 TTCCCAATTAAG 1962
 Db 2006 TTCCCAATTAAG 2017
 RESULT 12
 AF149783 1333 bp. mRNA linear PRI 02-JUL-2001
 LOCUS AF149783 Homo sapiens L-selectin ligand sulfotransferase mRNA, complete cds.
 DEFINITION AF149783
 ACCESSION AF149783
 VERSION AF149783.1 GI:13897503
 KEYWORDS
 SOURCE
 ORGANISM Homo sapiens (human)
 Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE 1 (bases 1 to 1333)
 Yeh, J.C., Hiraoaka, N., Petryniak, B., Nakayama, J., Ellies, L.G.,
 Rabuka, D., Hindsgraul, O., March, J.D., Lowe, J.B. and Fukuda, M.
 Novel sulfated lymphocyte homing receptors and their control by a
 Core1 extension beta 1,3-N-acetylglucosaminyltransferase
 Cell 105 (7), 957-969 (2001)
 JOURNAL MEDLINE 21332592
 MEDLINE 11439191
 PUBMED 2 (bases 1 to 1333)
 REFERENCE Hiraoaka, N. and Fukuda, M.
 AUTHORS Direct Submission
 TITLE Submitted (10-MAY-1999) Glycobiology, The Burnham Institute, 10901
 JOURNAL North Torrey Pines Road, La Jolla, CA 92037, USA
 FEATURES
 source 1. 1333

QY 814 GGGGAGCATGAGCAAAAACCTCAGAGAGAGGAGCAACCCCTACTATGTGATGAGTCAAGTCAAT 873
 DB 868 GGGGAGCATGAGCAAAAACCTCAGAGAGAGGAGCAACCCCTACTATGTGATGAGTCAAGTCAAT 927
 QY 874 CTGCCAAAAGCCAGCTGGAGATCTACAGACCATCTCAAGTCTTGGCCCAAGCCCTGCAGAGA 933
 DB 928 CTGCCAAAAGCCAGCTGGAGATCTACAGACCATCTCAAGTCTTGGCCCAAGCCCTGCAGAGA 987
 QY 934 ACCGTACCGCTTGTGGCTATGAGGACCTGGCTGAGGCCCTGTTGGGCCACAGACTTCCCG 993
 DB 988 ACCGTACCGCTTGTGGCTATGAGGACCTGGCTGAGGCCCTGTTGGGCCACAGACTTCCCG 1047
 QY 994 AATGATGAAATTCGTGGATGGAATTCCTTGGCCCATCTTCAAGACCTGGCTGATTAACAT 1053
 DB 1048 AATGATGAAATTCGTGGATGGAATTCCTTGGCCCATCTTCAAGACCTGGCTGATTAACAT 1107
 QY 1054 CACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1113
 DB 1108 CACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1167
 QY 1114 TGTCTCCAGAGCTTGGCGCTGCTTGGCCCTATGAAAAGTTTCTCGACTTCAGAAAC 1173
 DB 1168 TGTCTCCAGAGCTTGGCGCTGCTTGGCCCTATGAAAAGTTTCTCGACTTCAGAAAC 1227
 QY 1174 CTGTGGCGATGCGATGAATTTCTGCGGCTACCGCCACGTCAGATCTGAACAAGACAGAG 1233
 DB 1228 CTGTGGCGATGCGATGAATTTCTGCGGCTACCGCCACGTCAGATCTGAACAAGACAGAG 1287
 QY 1234 AATCGTGTGGATGCTCTGCTACCTGAGAGCTGCTGCTGAGCAAAATCCATAGAGG 1293
 DB 1288 AATCGTGTGGATGCTCTGCTACCTGAGAGCTGCTGCTGAGCAAAATCCATAGAGG 1347
 QY 1294 TTGAGAAAGCTTGTGCTGACCATCTGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1353
 DB 1348 TTGAGAAAGCTTGTGCTGACCATCTGCTGCTGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1407
 QY 1354 GCGTTCCTACATCTCTGAGCTTACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1413
 DB 1408 GCGTTCCTACATCTCTGAGCTTACATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1467
 QY 1414 TGTCTCCAGAGCTTGGCGCTGCTTGGCCCTATGAAAAGTTTCTCGACTTCAGAAAC 1473
 DB 1468 TGTCTCCAGAGCTTGGCGCTGCTTGGCCCTATGAAAAGTTTCTCGACTTCAGAAAC 1527
 QY 1474 ACTGGGGAACCTTATGTGAGAGCAGACATCCACAGTGAACAGGATTTGCTCTCTC 1533
 DB 1528 ACTGGGGAACCTTATGTGAGAGCAGACATCCACAGTGAACAGGATTTGCTCTCTC 1587
 QY 1534 TTTTCTGATCTTCTGCTGCTGAGAGCTTCAAGACTTGGCTGAGAGGCTATTA 1593
 DB 1588 TTTTCTGATCTTCTGCTGCTGAGAGCTTCAAGACTTGGCTGAGAGGCTATTA 1647
 QY 1594 GCACGACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1653
 DB 1648 GCACGACACAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1707
 QY 1654 GGGAAATGATCTTTCACCAAGAGCTCACACAGATTTTCACAGAGATCGCAATTTCTGAG 1713
 DB 1708 GGGAAATGATCTTTCACCAAGAGCTCACACAGATTTTCACAGAGATCGCAATTTCTGAG 1767
 QY 1774 CCCTTGGAGTTCGCAATGGATTCACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1773
 DB 1768 CCCTTGGAGTTCGCAATGGATTCACAGAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 1826
 QY 1774 ATGAGCTTGACCAT-ACAGCTATCGGTAATCAGAAATATGAAAACAAATCTGTG-ACAAA 1831
 DB 1827 ATGAGCTTGACCATCAGAGCTATCGGTAATCAGAAATATGAAAACAAATCTGTGCAAAA 1886
 QY 1832 AAGAGAGCTCTTAAGTTACAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1890
 DB 1887 AAGAGAGCTCTTAAGTTACAAAGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1946
 QY 1891 CATTTTCCATTAACATAGAAAACCTTGACCTGTGAACCTTGGCATCTGTTAATACTAANA 1950

DB 1947 CATTTTCCATCAGATAGAGAGCTTGACCTGAGAGC-TGCCATCTGTTAATACTAANA 2005
 QY 1951 TTCCCAATAG 1962
 DB 2006 TTCCCAATAG 2017
 RESULT 11
 AF131235
 LOCUS AF131235 2032 bp mRNA linear PRI 10-AUG-1999
 DEFINITION Homo sapiens N-acetylglucosamine 6-O-sulfotransferase mRNA,
 complete cds.
 ACCESSION AF131235
 VERSION AF131235.1 GI:4927113
 KEYWORDS
 SOURCE
 ORGANISM
 Homo sapiens (human)
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
 REFERENCE
 AUTHORS Bistrup, A., Bhakta, S., Lee, J. K., Below, Y. Y., Gunn, M. D., Zuo, F. R.,
 Huang, C. C., Kannagi, R., Rosen, S. D., and Hemmerich, S.
 TITLE Reconstitution of two specificities function in the
 JOURNAL J. Cell Biol. 145 (4), 899-910 (1999)
 MEDLINE 9926436
 PUBMED 10330415
 REFERENCE
 2 (bases 1 to 2032)
 Bistrup, A., Tangemann, K., Bhakta, S., Lee, J. K., Below, Y. Y.,
 Gunn, M. D., Zuo, F. R., Huang, C. C., Kannagi, R., Rosen, S. D., and
 Hemmerich, S.
 TITLE Direct Submision
 JOURNAL Submitted (25-FEB-1999) Respiratory Diseases, Roche Bioscience,
 3401 Hillview Avenue, Palo Alto, CA 94304, USA
 FEATURES
 source
 1..2032
 /organism="Homo sapiens"
 /mol_type="mRNA"
 /db_xref="taxon:9606"
 /cell_type="high endothelial"
 /tissue_type="tcons11"
 182..1342
 /function="biosynthesis of endothelial ligands for
 L-selectin"
 /note="HEC-GlcNAc-6-Sr"
 /codon_start=1
 /product="N-acetylglucosamine 6-O-sulfotransferase"
 /protein_id="AAD33015.1"
 /db_xref="GI:4927114"
 /translation="MLPKMKLLFLVSOMAILAFPHYSHNISLSMAKPERMH
 VLVSMSRSGSFFVGLFGQHPDVFILMEPAHVMYTFKOSTAMHLMHVRDLRAVE
 LDMVSFVDMERPGROSSLPOMENSRLGASPCDDIIPROEIPBRAHRLGSSOP
 FEVVERKASRSHVYLKFEVFNLSISVLPDLNLIHVLVLDPAVPRFSPRETK
 GLMIDSRIVKQHEKLEKEDQPIYVVOVITOSOLEIKTQSIPKALQERILVRY
 EDLARPVAQTSRMTVEVLEFLPHQTVNHNITRKGGHDAHFTNARDALNVSOAM
 RNSLPEKYSRLQKACGDAMNLLGRRHVSDEBORNLIDLILSTVPEQIH"
 BASE COUNT 468 a 569 c 490 g 505 t
 ORIGIN
 Query Match 87.4%; Score 1784.8; DB 9; Length 2032;
 Best Local Similarity 98.8%; Pred. No. 0;
 Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;
 QY 94 AAGCCCAACGACCAAGGCTTCCACTTACGACCAATGCTACTGCTAATAAATAATGAAGCT 153
 DB 148 AAGCCGCTTTCGAAAGCTTTCACCTTACGACCAATGCTACTGCTAATAAATAATGAAGCT 207
 QY 154 CCGTCTGTTTCTGCTTCCAGATGCGCATCTTGCGCTATCTTCCACATGTACAGCA 213
 DB 208 CCGTCTGTTTCTGCTTCCAGATGCGCATCTTGCGCTATCTTCCACATGTACAGCA 267
 QY 214 CAACATCAGCTCCCTCTCTATGAGGACACAGCCCGGCGCATGACAGCTGCTGCTCTGTC 273

```

Db 1228 CTGTGGGATGACATGATTTGCTGGGCTACCGCCACGTCATGATCTGACAGAACAGAG 1287
QY 1234 AACCTGTTCTGATGATCTGCTACCTGAGAGTCCCTGAGCAATTCACATAGAGGG 1293
Db 1288 AACCTGTTCTGATGATCTGCTACCTGAGAGTCCCTGAGCAATTCACATAGAGGG 1347
QY 1294 TTGAGAGAGCTTTGCTGACCACTGCTGTCACCTGACCTCACTTCTGATGCTTGA 1353
Db 1348 TTGAGAGAGCTTTGCTGACCACTGCTGTCACCTGACCTCACTTCTGATGCTTGA 1407
QY 1354 GCCTGCTGATCTGATCTGACCTTAACTACATGCTGTGGGTATCAGACAGCTGATG 1413
Db 1408 GCTTGTGATGATCTGATCTGACCTTAACTACATGCTGTGGGTATCAGACAGCTGATG 1467
QY 1414 TGTGTCCACAGCTGCTGACCACTGCTGTCACCTGACCTCACTTCTGATGCTTGA 1473
Db 1468 TGTGTCCACAGCTGCTGACCACTGCTGTCACCTGACCTCACTTCTGATGCTTGA 1527
QY 1474 ACTGGGACCTTATGTCAGACACATCCACAGTGAACAGAGTATTGCTTCTTC 1533
Db 1528 ACTGGGACCTTATGTCAGACACATCCACAGTGAACAGAGGATTGCTTCTTC 1587
QY 1534 TTTTCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1593
Db 1588 TTTTCTGATCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1647
QY 1594 GCACGACACAGTATGTCAGTGAATGATCATTAACCTGCTGCTGCTGCTGCTGCTGCT 1653
Db 1648 GCACGACACAGTATGTCAGTGAATGATCATTAACCTGCTGCTGCTGCTGCTGCTGCT 1707
QY 1654 GGAATGATCTTTCACCAAGAGCTCAGACGATTTTCCACAGAGATGCGAATTCGAG 1713
Db 1708 GGAATGATCTTTCACCAAGAGCTCAGACGATTTTCCACAGAGATGCGAATTCGAG 1767
QY 1714 CCTTGGAGTCCCAATGATTCAGAGGAAAGGAAAGGTTGATGCTGCTGCT 1773
Db 1768 CCTTGGAGTCCCAATGATTCAGAGGAAAGGAAAGGTTGATGCTGCTGCTGCT 1826
QY 1774 ATGAGCTTGACAT-ACAGCTATCGTATCAGAAATATGAAACAAATCTCTG-ACAA 1831
Db 1827 ATGAGCTTGACAT-ACAGCTATCGTATCAGAAATATGAAACAAATCTCTG-ACAA 1886
QY 1832 AGAGCAAGCTTATGTCAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1880
Db 1887 AGAGCAAGCTTATGTCAGAGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 1946
QY 1891 CATTTTCCATTCATAGAAACCTTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 1950
Db 1947 CATTTTCCATTCATAGAAACCTTGAAGTCTGCTGCTGCTGCTGCTGCTGCTGCT 2005
Y 1951 TTCCCAATTAAG 1962
Db 2006 TTCCCAATTAAG 2017

RESULT 10
BD134772 2032 bp DNA linear PAT 18-SEP-2002
LOCUS BD134772 Glycosylsulfoltransferase-3.
DEFINITION BD134772
ACCESSION BD134772
VERSION BD134772.1 GI:23229717
KEYWORDS JP 2002507409-A/1.
SOURCE Homo sapiens (human)
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistup,A., Rosen,S.D., Tangemann,K. and Hemmerlich,S.
TITLE Glycosylsulfoltransferase-3
JOURNAL Patent: JP 2002507409-A 12-MAR-2002;
THE REGENTS OF THE UNIVERSITY OF CALIFORNIA, SYNTAX INC
COMMENT OS Homo sapiens (human)
PN JP 2002507409-A/1

```

```

PD 12-MAR-2002
PF 26-FEB-1999 JP 2000537979
PR 20-MAR-1998 US 09/045284,12-NOV-1998 US 09/190911 PI
ANNEETTE BISTRUP, STEVEN D ROSEN, KIRSTEN TANGEMANN, STEFAN PI
HEMERLICH
PC C12N9/10,A01K67/027,A61K45/00,A61K48/00,A61P29/00,A61P37/06,
PC A61P43/00,
PC C12N1/15,C12N1/21,C12N5/10,C12N15/09,C12Q1/48,C12N5/00,C12N15/
PC 00
CC Glycosylsulfoltransferase-3
FH Key Location/Qualifiers
FT source 1..2032
FT 1..2032
FEATURES
source 1..2032
location/Qualifiers
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN
Query Match 87.4%; Score 1784.8; DB 6; Length 2032;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;
QY 94 AACCCAAAGCCACAGAGCTTCCACTTCACAGACATGCTACTGCTAAAAAATGAAGCT 153
Db 148 AAGCCGCTTTGCAAGAGCTTCCACTTCACAGACATGCTACTGCTAAAAAATGAAGCT 207
QY 154 CTGCTCTTTCTGCTTTCCAGATGCGCATCTTGGCTCTATCTTCCACATGACAGCA 213
Db 208 CCGTCTTTCTGCTTTCCAGATGCGCATCTTGGCTCTATCTTCCACATGACAGCA 267
QY 214 CAACATAGAGCTCCCTGCTATGAGAGCAGACGCCAGCAGATGAGTGGTGTGCTG 273
Db 268 CACATAGAGCTCCCTGCTATGAGAGCAGACGCCAGCAGATGAGTGGTGTGCTG 327
QY 274 TTCCTGCGCTTGGCTCTTCTTTTGGGGCAGCTTTTGGGAGCAGCCAGATGTTTT 333
Db 328 TTCCTGCGCTTGGCTCTTCTTTTGGGGCAGCTTTTGGGAGCAGCCAGATGTTTT 387
QY 334 CTACCTATGAGAGCCCGCTGCGCTGCTGATACCTTCAAGAGAGCAGCCGCTGAT 393
Db 388 CTACCTATGAGAGCCCGCTGCGCTGCTGATACCTTCAAGAGAGCAGCCGCTGAT 447
QY 394 GCTGACATGAGCTGCGGATGCTGATACGGGCGCTCTGCTGCGACATGAGCGTCT 453
Db 448 GCTGACATGAGCTGCGGATGCTGATACGGGCGCTCTGCTGCGACATGAGCGTCT 507
QY 454 TGATGCTATGAGAGCTGCTGCGGAGAGCAGTCCAGCTCTTTCAGTGGAGAAAG 513
Db 508 TGATGCTATGAGAGCTGCTGCGGAGAGCAGTCCAGCTCTTTCAGTGGAGAAAG 567
QY 514 CCGGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 573
Db 568 CCGGGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 627
QY 574 GGTGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 633
Db 628 GGTGCTGAGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 687
QY 634 CTCTACAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 693
Db 688 CTCTACAGCAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 747
QY 694 GCTGCTGAGAGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 753
Db 748 GCTGCTGAGAGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 807
QY 754 CGTGTCCCTTCCGAGAGAGGAGAGGAGATGATGATGATGATGATGATGATGATGATGAT 813
Db 808 CGTGTCCCTTCCGAGAGAGGAGAGGAGATGATGATGATGATGATGATGATGATGATGAT 867

```

DB 1399 ACCCTTGCCCTACATCTCTAGCCCTTACATACATGCTGTGGATATCAGCTAGCTGTAG 1458
QY 1413 TTGTGTCCACAGCTGTCTCAAGCAGAGAGACTTTTGTGTCCATGCTTGTGTCTAGAAACA 1472
DB 1459 TTGTGTCCACAGCTGTCTCAAGCAGAGAGACTTTTGTGTCCATGCTTGTGTCTAGAAACA 1518
QY 1473 GACTGGGGAACCTTATGTAGACAGACATCCACAGTGAACAGGGTATTGCTCTCTT 1532
DB 1519 GACTGGGGAACCTTATGTAGACAGACATCCACAGTGAACAGGGTATTGCTCTCTT 1578
QY 1533 CTTTCTTATGCTCTGCTGTGGGAGACTTCAGAGACTTTGGCTGGAGGCCATTAA 1592
DB 1579 CTTTCTTATGCTCTGCTGTGGGAGACTTCAGAGACTTTGGCTGGAGGCCATTAA 1638
QY 1593 AGCAGCAGACAGTATCAGTGAATGTATCAATCAACCTCCCTGTCCACATCTTGGCCAA 1652
DB 1639 AGCAGCAGACAGTATCAGTGAATGTATCAATCAACCTCCCTGTCCACATCTTGGCCAA 1698
QY 1653 GGGGAATGATCTTTCACCAAGAGCTCACAGCATTTTCCACAGATGCCAATTCTGA 1712
DB 1699 GGGGAATGATCTTTCACCAAGAGCTCACAGCATTTTCCACAGATGCCAATTCTGA 1758
QY 1713 GCCCTTGGATCTCCCATGGGATTCAGAGAGAGAGTGGGAACAAGTGGATGCCCTACT 1772
DB 1759 GCCCTTGGATCTCCCATGGGATTCAGAGAGAGAGTGGGAACAAGTGGATGCCCTACT 1817
QY 1773 TATGAGCTTGACCAT-ACAGCTATCGGTATCAGAAATATGAACAAATCTCTG-ACAA 1830
DB 1818 TATGAGCTTGACCATCAGCTATCGGTATCAGAAATATGAACAAATCTCTGACAA 1877
QY 1831 AAGAGCAGCTTATAGTTTACCAAGTGGCTGGCTGTGATTTGATATCA-TTCCCTTT 1889
DB 1878 AAGAGCAGCTTATAGTTTACCAAGTGGCTGGCTGTGATTTGATATCA-TTCCCTTT 1937
QY 1890 GCATTTTCCATTAACATAAACTTGACCTGTGAACTTGACCTGTGATTAATATA 1949
DB 1938 GCATTTTCCATTAACATAAACTTGACCTGTGAACTTGACCTGTGATTAATATA 1996
QY 1950 ATTCCCAATAAGGTTCTGTATAGAAATGCCCTTTTATGCTTCTTAATTATAGCAGTA 2009
DB 1997 ATTCCCAATAAGGTTCTGTATAGAAATGCCCTTTTATGCTTCTTAATTATAGCAGTA 2056
QY 2010 AATGTTTCA 2018
DB 2057 AATGTTTCA 2065

RESULT 9
AR203335
LOCUS AR203335 2032 bp DNA linear PART 20-JUN-2002
DEFINITION Sequence 2 from patent US 6365365.
ACCESSION AR203335
VERSION AR203335.1 GI:21499698
KEYWORDS
SOURCE Unknown.
ORGANISM Unknown.
REFERENCE 1 (bases 1 to 2032)
AUTHORS Bistup,A., Rosen,S.D., Tangemann,K. and Hemmerich,S.
TITLE Method of determining whether an agent modulates glycosyl
JOURNAL Patent: US 6365365-A 2 02-APR-2002;
FEATURES
source Location/Qualifiers
BASE COUNT 468 a 569 c 490 g 505 t
ORIGIN
Query Match 87.4%; Score 1784.8; DB 6; Length 2032;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;
QY 94 AAGCCCAAGCCACAGGCTTCCACTTCAGCACAATGCTACTGCTTAAATAATGAGCT 153

DB 148 AAGCCCGCTTCCAGAGCTTCCACTTCAGCACAATGCTACTGCTTAAATAATGAGCT 207
QY 154 CCGGCTTTTCTGTTTCCAGATGGCAGCAATCTGGCTATCTTCCACATGACAGCA 213
DB 208 CCGGCTTTTCTGTTTCCAGATGGCAGCAATCTGGCTATCTTCCACATGACAGCA 267
QY 214 CAACATCAGCTCCCTGTATGAAAGCCAGACGCCAGCCATGACAGCTGCTGTCTC 273
DB 268 CAACATCAGCTCCCTGTATGAAAGCCAGACGCCAGCCATGACAGCTGCTGTCTC 327
QY 274 TTCTGGGCTCTGGCTCTTCTTTTGTGGGCAAGCTTTTGGGCAAGCAGATTTT 333
DB 328 TTCTGGGCTCTGGCTCTTCTTTTGTGGGCAAGCTTTTGGGCAAGCAGATTTT 387
QY 334 CTACCTGATGAGAGCCCGCTGGCAGCTGTGATGATCAATCAAGCAGACAGCCCTGAT 393
DB 388 CTACCTGATGAGAGCCCGCTGGCAGCTGTGATGATCAATCAAGCAGACAGCCCTGAT 447
QY 394 GCTGCATGAGCTGTGCGGGATCTGATACGGGCGCTTCTTGTGCGATGAGCGCTT 453
DB 448 GCTGCATGAGCTGTGCGGGATCTGATACGGGCGCTTCTTGTGCGATGAGCGCTT 507
QY 454 TGATGCTTACATGGAACCTGTGTCCCGGAGACAGTCCAGCTTTTCAATGGAGAACG 513
DB 508 TGATGCTTACATGGAACCTGTGTCCCGGAGACAGTCCAGCTTTTCAATGGAGAACG 567
QY 514 CCGGCGCTGTGTGTCACCTGCTGTGATCAATCAATCAAGCAGATGGAATCAATCCCGG 573
DB 568 CCGGCGCTGTGTGTGTCACCTGCTGTGATCAATCAATCAAGCAGATGGAATCAATCCCGG 627
QY 574 GCGTCACTCAGGCTCTGTGACAGTCAAGCAGCTTGTGAGTGTGAGAGAGCGCTCGG 633
DB 628 GCGTCACTCAGGCTCTGTGACAGTCAAGCAGCTTGTGAGTGTGAGAGAGCGCTCGG 687
QY 634 CTCTTACAGCAGCTGTGTCTCAAGAGAGTGGCTTCTTCAACTGCACTGCTTCAAGC 693
DB 688 CTCTTACAGCAGCTGTGTCTCAAGAGAGTGGCTTCTTCAACTGCACTGCTTCAAGC 747
QY 694 GCGTCAAGAGAGCCCTGCTCAACTGATATCGACAGCTGTGCGGAGACCGCCGGG 753
DB 748 GCGTCAAGAGAGCCCTGCTCAACTGATATCGACAGCTGTGCGGAGACCGCCGGG 807
QY 754 CGTGTTCCTGTCGCGAGAGAGCAAGAGGAGATCTCATGATGACAGTGCATTTGAT 813
DB 808 CGTGTTCCTGTCGCGAGAGAGCAAGAGGAGATCTCATGATGACAGTGCATTTGAT 867
QY 814 GGGGCGCATGAGCAAAAATCAAGAGAGAGAGCAACCTTATGATGAGTCAAT 873
DB 868 GGGGCGCATGAGCAAAAATCAAGAGAGAGAGCAACCTTATGATGAGTCAAT 927
QY 874 CTGCGAAGAGAGCTGGAGATCTACAAAGACATCCAGTCTTGGCCCAAGGCCCTGAGGA 933
DB 928 CTGCGAAGAGAGCTGGAGATCTACAAAGACATCCAGTCTTGGCCCAAGGCCCTGAGGA 987
QY 934 ACGTACCTGCTTGTGCGCTATGAGAGACCTGTGCGAGCCCTGTGCGAGACTTCCG 993
DB 988 ACGTACCTGCTTGTGCGCTATGAGAGACCTGTGCGAGCCCTGTGCGAGACTTCCG 1047
QY 994 AATGTATGATTCGTGGGATTTGGAATCTTGGCCCATCTTTCAGAGACTGGGTGATTAAT 1053
DB 1048 AATGTATGATTCGTGGGATTTGGAATCTTGGCCCATCTTTCAGAGACTGGGTGATTAAT 1107
QY 1054 CACCGAGAGAGAGGAGTGGTACACAGCTTTCACACAAATGCCAGGATGCCCTTAA 1113
DB 1108 CACCGAGAGAGAGGAGTGGTACACAGCTTTCACACAAATGCCAGGATGCCCTTAA 1167
QY 1114 TGTCTCCAGGCTTGGCGCTGTCTTGTGCTTATGAAAGTTCTGCACTTCAGAACG 1173
DB 1168 TGTCTCCAGGCTTGGCGCTGTCTTGTGCTTATGAAAGTTCTGCACTTCAGAACG 1227
QY 1174 CTGTGGGATGCGATGAAATTTCTGTGGCTACCGGCGCATGCTGAAACAAGACAGG 1233

```

Db 123928 CCTGTGGGAGACCTTCAGACATTTGTGGCGCTGGAGCCTATTAGCAGACAGTA 123869
Qy 1607 TCAGTGAATGATCCATAACCTCCCTGTCCACATCTTGGCCAAATGGGATGATCT 1666
Db 123868 TCAGTGAATGATCCATAACCTCCCTGTCCACATCTTGGCCAAATGGGATGATCT 123809
Qy 1667 TCACCAAGAGCTCACCAGCATTTTCCACAGAGATGCGAATTTGAGCCCTTGGAGTCC 1726
Db 123808 TCACCAAGAGCTCACCAGCATTTTCCACAGAGATGCGAATTTGAGCCCTTGGAGTCC 123749
Qy 1727 CAATGGGATTCAGAGGAGAGAGTGGGAACAGGTTGGATGCCCTACTTTAGAGCTTGACCA 1786
Db 123748 CAGT-GGATTCAGAGGAGAGAGTGGGAACAGGTTGGATGCCCTACTTTAGAGCTTGACCA 123690
Qy 1787 T-ACAGTATGCGTATATGAGAAATATGAAACAAAATCTCTG-ACAAAAGAGACAGCTTT 1844
Db 123689 TCACAGCATATGGTATATGAGAAATATGAAACAAAATCTCTGACAAAAGAGACAGCTTT 123630
Qy 1845 AAGTTCACAGAGTGCCTGGGCTGTGATTTGAATATCA-TTTCCTTTGGCATTTTCCCATTA 1903
Db 123629 AAGTTCACAGAGTGCCTGGGCTGTGATTTGAATATCA-TTTCCTTTGGCATTTTCCCATTA 123570
Qy 1904 CATGAAAACCTTTGACCTGTGTAACCTTCCATCTGTATTAATTAATTTCCCAATAGG 1963
Db 123569 CATGAAAACCTTTGACCTGTGTAACCTTCCATCTGTATTAATTAATTTCCCAATAGG 123511
Qy 1964 TTCTGTTTGAATGATCCCTTTTATAGCTTTTAAATTTATAGCAGTAATGTTCATTTTA 2023
Db 123510 TTCTGTTTGAATGATCCCTTTTATAGCTTTTAAATTTATAGCAGTAATGTTCATTTTA 123451
Qy 2024 TGGGATCCTAAAAAAA 2042
Db 123450 TGGGATCCTAAGCTAAGNA 123432

RESULT 8
AR310485
LOCUS AR310485 2065 bp DNA linear PAT 12-JUN-2003
DEFINITION Sequence 26 from patent US 6558935.
ACCESSION AR310485
VERSION AR310485.1 GI:31703448
KEYWORDS
SOURCE
ORGANISM
Unknown.
Unclassified.
1 (bases 1 to 2065)
AUTHORS Tang,X.T., Corley,N.C., Guegler,K.J., Baughn,M.R., Lal,P., Yue,H.,
Hillman,J.L. and Azimzal,Y.
Human transferrase proteins
Patent: US 6558935-A 26 06-MAY-2003;
JOURNAL Location/Qualifiers
FEATURES
source
1..2065
/organism="unknown"
BASE COUNT 466 a 575 c 491 g 533 t
ORIGIN
Query Match 88.7%; Score 1813; DB 6; Length 2065;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1902; Conservative 0; Mismatches 20; Indels 7; Gaps 7;

```

```

Db 320 TTCTGGGCGCTGTGCTCTTCTTTGTGGGAGAGCTTTTGGGAGACCCAGATGTTT 379
Qy 334 CTACCTGATGAGACCCGCTGTGACGTGTGATGACCTTCAAGACGAGACCCGCTGAT 393
Db 380 CTACCTGATGAGACCCGCTGTGACGTGTGATGACCTTCAAGACGAGACCCGCTGAT 439
Qy 394 GCTGACATGAGCTGTGGGAGATCTGATATGAGGCGCTCTTCTTGTGCGACATGAGCGT 453
Db 440 GCTGACATGAGCTGTGGGAGATCTGATATGAGGCGCTCTTCTTGTGCGACATGAGCGT 499
Qy 454 TGATGCTACATGAGAACCTGTCCCGGAGACAGTCAGAGCTTTTCAAGTGGAGAGACAG 513
Db 500 TGATGCTACATGAGAACCTGTCCCGGAGAGACAGTCAGAGCTTTTCAAGTGGAGAGACAG 559
Qy 514 CCGGGCCCTGTGTTGTGACCTGCTGTGACATCATCCCAAGATGAAATCATCCCCG 573
Db 560 CCGGGCCCTGTGTTGTGACCTGCTGTGACATCATCCCAAGATGAAATCATCCCCG 618
Qy 574 GGCCTACAGAGCTCCTGTGACAGTCAACAGAGCTTTGAGGTGG-TGGAGAGAGCTGCC 632
Db 619 GGCCTACAGAGCTCCTGTGACAGTCAACAGAGCTTTGAGGTGGAGAGAGCTGCC 678
Qy 633 GCTCCTACAGAGCTGCTGTGCAAGAGAGTGGCTTCTCAACCTGACAGTCCCTAC 692
Db 679 GCTCCTACAGAGCTGCTGTGCAAGAGAGTGGCTTCTCAACCTGACAGTCCCTAC 738
Qy 693 CGCTGCTAAGAGACCCCTCCCTCAACCTGCAATGCTGACCTGCTCCGGAGCCCGGG 752
Db 739 CGCTGCTAAGAGACCCCTCCCTCAACCTGCAATGCTGACCTGCTCCGGAGCCCGGG 798
Qy 753 CGCTGCTCCTGTCCCGAAGACGACAAAGGAGATCTGATTTGACAGTGGCATTTGA 812
Db 799 CGCTGCTCCTGTCCCGAAGACGACAAAGGAGATCTGATTTGACAGTGGCATTTGA 858
Qy 813 TGGGCGACATGAGCAAAACTCAAGAGAGAGACCAACCTTATGTATGACAGTCA 872
Db 859 TGGGCGACATGAGCAAAACTCAAGAGAGAGAGACCAACCTTATGTATGACAGTCA 918
Qy 873 TCTGCAAGAGCAGTGGAGATCTCAAGACATCCATCTTGGCCAAAGCCCTGCAGG 932
Db 919 TCTGCAAGAGCAGTGGAGATCTCAAGACATCCATCTTGGCCAAAGCCCTGCAGG 978
Qy 933 AAGCTACCTGCTGTGCTATGAGAGCTGTGCTGAGCCCTGTGGCCAGACTTCC 992
Db 979 AAGCTACCTGCTGTGCTATGAGAGCTGTGCTGAGCCCTGTGGCCAGACTTCC 1038
Qy 993 GAATGATGAATTTGTTGGATTTGGAATTTTCCCATCTTCAAGACCTGGGTGCTATCA 1052
Db 1039 GAATGATGAATTTGTTGGATTTGGAATTTTCCCATCTTCAAGACCTGGGTGCTATCA 1098
Qy 1053 TCACCGAGAGCAGAGGATGGGTGACACAGCTTTCACACAAATCCAGAGGTGCCCTTA 1112
Db 1099 TCACCGAGAGCAGAGGATGGGTGACACAGCTTTCACACAAATCCAGAGGTGCCCTTA 1158
Qy 1113 ATGCTCCAGAGCTTGGGCTGTGCTTTCCTTGAAGAGTTTCTGACTTGAAAG 1172
Db 1159 ATGCTCCAGAGCTTGGGCTGTGCTTTCCTTGAAGAGTTTCTGACTTGAAAG 1218
Qy 1173 CCTGTGGGAGATGATGATTTTGGGCTTACCGGACAGTGTGATCTGAATCAAGACGA 1232
Db 1219 CCTGTGGGAGATGATGATTTTGGGCTTACCGGACAGTGTGATCTGAATCAAGACGA 1278
Qy 1233 GAAACCTGTGCTGATCTTGTGCTTACCTGAGCTGCTGAGAGAAATTCACATAGAG 1292
Db 1279 GAAACCTGTGCTGATCTTGTGCTTACCTGAGCTGCTGAGAGAAATTCACATAGAG 1338
Qy 1293 GTTGAAGAGCTTGTGCTCAGCTGTGCTCAGCTCAGCTCACTTCTCTGATGCTTCTG 1352
Db 1339 GTTGAAGAGCTTGTGCTCAGCTGTGCTCAGCTCAGCTCACTTCTCTGATGCTTCTG 1398
Qy 1353 AGCTTGTGCTCAGCTGTGAGCTTAACTACATGCTGTGGGTATCACATGAGTGTAG 1412

```

THIS PAGE BLANK (USPTO)

```

Db      1172  GCGTTCATCTTCGCTAGGAATGCGGCAACGCTCCAGGCTGCGCCACGCGFTG 1231
QY      1142  CCCATGAAAGGTTTCTCGACTCAGAAAGCCGTGCGCATGCCATGAATTGCTGSGC 1201
Db      1232  CCTTCACTAGATCTGCGGCTGAGAGGTGTGCGCGGCGCTGCAGCTGCTGSGC 1291
QY      1202  TACCGCCACGTCAAGATCTGAACAAGAACGAGAAACCTGTGCTGATCTTCTG 1255
Db      1292  TACCGGCTGTGTACTCTGCGGACACGAGCGGTGACCTCACCTGATCTGTG 1345

```

Search completed: August 14, 2003, 11:38:20
 Job time : 514.533 secs

FT	/tag= h
FT	/cons_splice= (5'site:YES, 3'site:NO)
FT	47939..49746
FT	/tag= i
FT	number= 5
FT	/note= "Includes 17 base pairs of 5'UTR, the ORF
FT	and all of 3'UTR"
FT	47939..47955
FT	/tag= j
FT	/note= "Portion of 5' untranslated region (5'UTR)"
FT	47956..49128
FT	/tag= k
FT	/product= "Human glycosyl transferase-4alpha
FT	(GST-4alpha)"
FT	49129..49746
FT	/tag= l
FT	83257..83347
FT	/tag= m
FT	/label= 4a_5U2
FT	83348..96412
FT	/tag= n
FT	/cons_splice= (5'site:NO, 3'site:NO)
FT	96413..96484
FT	/tag= o
FT	/label= 4a_5U1
FT	96485..98456
FT	/tag= p
FT	/cons_splice= (5'site:NO, 3'site:NO)
FT	98457..99968
FT	/tag= q
FT	/note= "Includes 17 base pairs of 5'UTR, the ORF
FT	and all of 3'UTR"
FT	98457..98473
FT	/tag= r
FT	/note= "Portion of 5' untranslated region (5'UTR)"
FT	98474..99661
FT	/tag= s
FT	/product= "Human glycosyl transferase-4beta
FT	(GST-4beta)"
FT	99662..99968
FT	/tag= t
PN	MO200106015-AI.
PD	25-JAN-2001.
PF	19-JUL-2000; 2000WO-US19741.
PR	20-JUL-1999; 99US-0144694.
PR	13-JUL-2000; 2000US-0593826.
RA	(RESC) UNIV CALIFORNIA.
PI	Rosen SD, Lee JK, Hemmerich S;
PJ	WPI: 2001-138471/14.
DR	P-PSDB: AAY72639, AAY72640.
XX	New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
XX	diagnostic and therapeutic agent screening applications -
XX	Example 1; Page 62-104; 128pp; English.
XX	The present sequence is human glycosyl sulfotransferase-4 (GSR-4) genomic
CC	DNA encoding GSR-4alpha and GSR-4beta. GSR-4 gene is found on
CC	chromosome 16q23.1.
CC	GSR is a type 2 membrane protein useful for inhibiting a binding event
CC	between a selectin and a selectin ligand, which comprises contacting the
CC	selectin with a non-sulphated selectin ligand, GSR and a small molecular
CC	agent that inhibits the sulphation activity of GSR. GSR is also useful
CC	in inhibiting a selectin mediated binding event. GSR is useful in gene
CC	therapy to treat disorders such as acute or chronic inflammation,
CC	systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis


```

FT  /tag- c
XX  W0200106015-A1.
XX  25-JAN-2001.
XX  19-JUL-2000; 2000MO-US19741.
XX  20-JUL-1999; 99US-0144694.
PR  13-JUL-2000; 2000US-0593828.
XX  (RECC ) UNIV CALIFORNIA.
XX  Rosen SD, Lee JK, Hemmerich S;
PI  WPI: 2001-138471/14.
DR  P-PSDB; AAY72640.
XX  New glycosyl sulfotransferases (GST)-4alpha, GST-4beta and GST-6 for
PT  diagnostic and therapeutic agent screening applications
XX  Claim 6; Fig 4A; 128pp; English.
XX  The present sequence is human glycosyl sulfotransferase-4beta (GST-4
CC  beta) cDNA. GST-4 gene is found on chromosome 16q23.1.
CC  GST is a type 2 membrane protein useful for inhibiting a binding event
CC  between a selectin and a selectin ligand, which comprises contacting the
CC  selectin with a non-sulphated selectin ligand, GST and a small molecular
CC  agent that inhibits the sulphation activity of GST. GST is also useful
CC  in inhibiting a selectin mediated binding event. GST is useful in gene
CC  therapy to treat disorders such as acute or chronic inflammation,
CC  systemic lupus erythematosus (SLE), rheumatoid arthritis, polyarteritis
CC  nodosa, polymyositis, dermatomyositis, systemic sclerosis, diabetes,
CC  glomerulonephritis, myasthenia gravis, Sjogren's syndrome, Hashimoto's
CC  disease, Grave's disease, adenitis, hypoparathyroidism, pernicious
CC  anemia, demyelinating diseases, cirrhosis, ulcerative colitis,
CC  dermatitis, myocarditis, regional enteritis, adult respiratory distress
CC  syndrome, infantile eczema, psoriasis lichen planus, allergic rhinitis,
CC  bronchial asthma, hypersensitivity, rheumatic fever and tissue rejection
CC  during transplantation.
XX  Sequence 1694 BP; 259 A; 586 C; 514 G; 335 T; 0 other;
SQ
Query Match 20.28; Score 412.8; DB 22; Length 1694;
Best Local Similarity 64.48; Pred. No. 1.2e-110;
Matches 653; Conservative 0; Mismatches 352; Indels 9; Gaps 2;
QY 248 GAGGGGATGACGAGTGTCTGTCTCTCCGCGGCTGCGCTCTTTGTGGGCGAG 307
DB 303 GCGGCGCTGCAATGCTGTGCTGTCTCTCGGCGCTGCGCTGCTCTGCGGCGAA 362
QY 308 CTTTTTGGGAGCACCCAGATGTTTCTACCTGATGAGCCCGCTGTGACAGTGTGATG 367
DB 363 CTCTTCAACGAGCAGCCCGAGAGTCTTACTTATGAGACCGCGGTGACAGTGTGAC 422
QY 368 ACCTTCAAGAGAGCAGCCCGTGTGATCTGCAATGCTGTGCGGATCTGATACGGGCC 427
DB 423 ACCCTGTGCGAGGCGAGCGCGCAACGCTGACATGCTGTGCGGAGCTGTGCGCTCC 482
QY 428 GTCTCTTGTGGCAGATGAGCGTGTGATGCTTACATGAGAACTGTGCTCCCGAGACAG 487
DB 483 GTCTCTCTGTGCGAGATGAGCGTGTGATGCTTATGTG---CCTTGGCGCGCAACTG 539
QY 488 TCCAGCCTCTTTCATGGAGAACAGCCGGCCCTGTGTTCTGACAGCTGCTGTGACATC 547
DB 540 TCCGACCTCTTTCATGGAGCGCGGTGAGCGTGTGCTGTGCGACCCGCGCTGAGTCC 599
QY 548 ATCCACAGATGAATATCATCCCGGGCTCATCTGACAGGCTCTGTGCAATCAACAGCCC 607
DB 600 TTTCGCCAGAGCGCATCACAGAGGCGGTGTGCAAGCACTGTGCGCGCGCAGTCC 659
QY 608 TTTCAGGTGTGAGAGAGCGCTGCGCTCTACAGCCAGTGTGTCTCAAGAGAGTCCGC 667

```

```

DB 660 TTCACCTGCGCCGGAGAGCCTCGCTCTACAGCCAGCTGTGTCTCAAGAGTCCGC 719
QY 668 TTCTTCAACCTGACGTCCCTTACCCGCTGTGTAAGACCCCTCCCTCAACCTCATATC 727
DB 720 TTCTTCAACCTGACGTCCCTTACCCGCTGTGTAAGACCCCTCCCTCAACCTCATATC 779
QY 728 GTGCACTGTGTCCGAGCCCGCGGCGCTGTGTCCTGTCCTGAGAACGACCAAGGAGAT 787
DB 780 GTGCACTGTGTCCGAGCCCGCGGCGCTGTGTCCTGTCCTGAGAACGACCAAGGAGAT 839
QY 788 CTCAATGATGACATGCTGATGTGATGAGGAGCATGACAAAATCTCAAGAGAGAC 847
DB 840 CTGCGCGGTGACCAAGGAGATGCTGTGAGCAACAGCGACAGTGGGTGAGACCCC 899
QY 848 CAACCTTACTATGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 907
DB 900 GCGCTGCGCGTGTGCGGAGAGTGTGCTGTGAGCAACCTGATGATGATGATGATGATG 959
QY 908 CAGTCTTGTCCCAAGGCGCTGCAAGAACGCTTACCTGTGTCCTTATGAGACCTGTGCT 967
DB 960 CTCAAGCGCGCACCTTCTGCGGCGCGCTTACCGCGCTGTGAGTGTGAGTGTGAGTGTG 1019
QY 968 CGAGCCCTGTGCGCCAGACTTCCGATGTATGATGTATGATGTGATGTGATGTGATG 1027
DB 1020 CGGAGCGCGTGTGAGAAATCCGTGCGCTTACCGCTTACCGCTTACCGCTTACCGCT 1079
QY 1028 CATCTTACAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1081
DB 1080 CAGCTGAGCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1139
QY 1082 GCTTTCACACAAATGCCAGGATGCCCTTATGTCTTCCAGGCTTGTGCTGTGCTTGTG 1141
DB 1140 GCTTTCACAAATGCCAGGATGCCCTTATGTCTTCCAGGCTTGTGCTGTGCTTGTG 1199
QY 1142 CCCATGAAAGTTTCTGCTGCTTCAAGAACCCCTGCGGCTGATGATGATGATGATGATG 1201
DB 1200 CCTTGTGCAAGATGCCCGCGGTGAGAACTGTGCGGTGAGTGTGAGTGTGAGTGTGAG 1259
QY 1202 TACCGCAGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1295
DB 1260 TACCGCCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 1313

```

RESULT 13
ABN89506 standard; cDNA: 2544 BP.

ABN89506:
05-SEP-2002 (first entry)

Human corneal N-acetylglucosamine-6-sulfotransferase cDNA seq ID NO:1.

Human: N-acetylglucosamine-6-sulfotransferase; enzyme: GLCNAC6ST;
corneal; sulfation; keratan sulfate; macular corneal dystrophy; MCD;
ophthalmological; chromosome 16q22; gene: ss.

Homo sapiens.

key Location/Qualifiers
CDS 693..1880
/tag- a
/product= "N-acetylglucosamine-6-sulfotransferase"

US2002061562-A1.
23-MAY-2002.
09-AUG-2001; 2001US-0927602.
11-AUG-2000; 2000US-325773P.
(FUKU/) FUKUDA M N.


```

Db 1576 ATCCACAGATGAATATCATCCCGGGCTCAGCTCAGGCTCTGTGAGTCACAGACCC 1635
Qy 608 TTGAGTGTGTGAGAGAGCCCTGCGCTCTCTCAGCAGCAGGTGTGTAAGAGGAGCC 667
Db 1636 TTGAGTGTGTGAGAGAGCCCTGCGCTCTCTCAGCAGCAGGTGTGTAAGAGGAGCC 1695
Qy 668 TTCTTCACCTGACGTCTCCCTTACCCGCTGTGAAGAACCCCTCCCTCAACCTGCATATC 727
Db 1696 TTCTTCACCTGACGTCTCCCTTACCCGCTGTGAAGAACCCCTCCCTCAACCTGCATATC 1755
Qy 728 GTGACACGTGTGAGAGAGCCCGGGGCGGTCTCCCTCCGAGAACGCAAGAGAGAT 787
Db 1756 GTGACACGTGTGAGAGAGCCCGGGGCGGTCTCCCTCCGAGAGAGAGAGAGAT 1815
Qy 788 CTCAATGATGACAGTGTGATGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 847
Db 1816 CTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1875
Qy 848 CAACCTCTACTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 907
Db 1876 CACCTGCGCGGTGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1935
Qy 908 CAGTCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 967
Db 1936 CTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1995
Qy 968 CGAGCCCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1027
Db 1996 CGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2055
Qy 1028 CATCTTCAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1081
Db 2056 CAGCTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2115
Qy 1082 GCTTTCACACAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1141
Db 2116 GCTTTCACACAAATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2175
Qy 1142 CCGTATGAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1201
Db 2176 CCGTATGAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2235
Qy 1202 TACCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1255
Db 2236 TACCGCAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 2289

```

RESULT 9
ABK54724
ID ABK54724 standard; cDNA; 517 BP.

```

AC ABK54724;
DT 18-JUN-2002 (first entry)
DE Human colon cancer-associated cDNA, SEQ ID No 194.
KW Human; colon cancer; immunogenic; vaccine; tumour; gene; ss.
XX Homo sapiens.
XX WO200212280-A2.
XX 14-FEB-2002.
XX 30-JUL-2001; 2001WO-US23826.
XX 03-AUG-2000; 2000US-223265P.
XX 02-OCT-2000; 2000US-237406P.
XX 20-MAR-2001; 2001US-277495P.
XX 03-JUL-2001; 2001US-302702P.
XX (CORI-) CORIXA CORP.

```

```

XX Pyle RA, Xu J, Secret H;
XX WPI; 2002-257462/30.
XX Novel polynucleotide encoding colon tumour polypeptides, useful as
XX vaccines for treating colon cancers
XX Claim 1; Page 206; 425pp; English.
XX The invention relates to isolated polynucleotides (I) encoding colon
XX tumour polypeptides (II). (I) is useful for stimulating an immune
XX response in a patient and treating colon cancer in a patient.
XX Oligonucleotides derived from (I) are useful for determining the presence
XX of cancer in a patient. (I) and (II) are useful in pharmaceutical
XX compositions, e.g. vaccines, and other compositions for the diagnosis
XX and treatment of colon cancer. A composition comprising a first component
XX selected from physiologically acceptable carriers and immunostimulants,
XX and an antigen-presenting cell expressing (II) is useful for inhibiting
XX development of cancer in a patient. (I) is useful in the design and
XX preparation of ribozyme molecules for inhibiting expression of tumour
XX polypeptides and (I). ABK54531-ABK55464 represent human colon cancer cDNA
XX sequences of the invention.
SQ Sequence 517 BP; 121 A; 130 C; 124 G; 142 T; 0 other;
Query Match 23.5%; Score 480.2; DB 24; Length 517;
Best local similarity 99.0%; Pred. No. 7.8e-131;
Matches 504; Conservative 0; Mismatches 3; Indels 2; Gaps 2;
Qy 1298 GAAGGCTTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1357
Db 10 GAGGCGTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 69
Qy 1358 TGCCTACATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1417
Db 70 TGCCTACATCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 129
Qy 1418 TCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1477
Db 130 TCCACAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 189
Qy 1478 GGAACCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1537
Db 190 GGAACCTTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 249
Qy 1538 CTTGATCTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1597
Db 250 CTTGATCTTCTGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 309
Qy 1598 GACACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1657
Db 310 GACACAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 369
Qy 1658 ATGATCTTTCACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1717
Db 370 ATGATCTTTCACAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 429
Qy 1718 TGGAGTCTCCATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1777
Db 430 TGGAGTCTCCATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 488
Qy 1778 GCTTGACCAT-ACAGTATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1805
Db 489 GCTTGACCATGAGATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 517

```

RESULT 10
ABV89280
ID ABV89280 standard; cDNA; 505 BP.
XX ABV89280;
XX

QY 711 CCTCAACCTGATATGCTGACCTGGTCCGGGACCCCGGGCGTGTCCGTCGGAG 770
 Db 985 CCTCAACCTGATATGCTGACCTGGTCCGGGACCCCGGGCGTGTCCGTCGGAG 1044
 QY 771 AAGCACAAGAGGAGATCTATGATGACANTGCCATTTGATGGGAGATGACGAAA 830
 Db 1045 AGCAGACACACCATGAACTCATGATGACATCTATGCTGATGGGAGAGATTGGAAA 1104
 QY 831 AACTCAAGAAAGAGAGACCAACCCCTACTATGATGACAGTGTGCGCAAGCCAGCTGG 890
 Db 1105 CGATGCAAGAGAGAGAGACAGCCCTATGATGACAGATGATCATGTGCAAAAGCCAGGTTG 1184
 QY 891 AGATCTCAAGAGACCATCTAGTCTGTCGCAAGGCCCTGACAGAAAGCTACCTGCTGTGC 950
 Db 1165 ACATAGTCAAGAGGACCATCTAGTCTGTCGCAAGGCCCTGACAGAAAGCTACCTGCTGTGC 1224
 QY 951 GCTATGAGAGACCTGGCTGAGCCCTGTGGCCGACCTGCGCAATGATGATATTCGTTGG 1010
 Db 1225 GGTATGAGAGACCTGGCTGAGCCCTGTGGCCGACCTGCGCAATGATGATATTCGTTGG 1284
 QY 1011 GATTGATATCTTGGCCCATCTTCCAGACCTGGTGCATTAACATCAGCCGAGGCAAGGGA 1070
 Db 1285 GATTGATATCTTGGCCCATCTTCCAGACCTGGTGCATTAACATCAGCCGAGGCAAGGGA 1344
 QY 1071 TGGGTGACACAGCTTTCACACAAATGCGCAGAGTGCCTTAATGCTCCAGGCTTGGC 1130
 Db 1345 TGGGTGACACAGCTTTCACACAAATGCGCAGAGTGCCTTAATGCTCCAGGCTTGGC 1404
 QY 1131 GCTGCTTTCGCTTATGAAAGGTTCTGACCTTCAAGAAAGCTGTGGGCTGATGCA 1180
 Db 1405 GTTGTGCTTTCGCTTATGAAAGGTTCTGACCTTCAAGAAAGCTGTGGGCTGATGCA 1464
 QY 1191 ATTGTGCGGCTACCGGACCTGATCTGAAAGAAAGAAAGAAAGCTGTGGTGCATG 1250
 Db 1465 ATTGTGCGGCTACCGGACCTGATCTGAAAGAAAGAAAGAAAGCTGTGGTGCATG 1524
 QY 1251 TTCTGCTACCTGAGCTGCTCCGAGCAATATCACTAAGAGGTTGAGAAAGCTTGTG 1310
 Db 1525 TTCTGCTACCTGAGCTGCTCCGAGCAATATCACTAAGAGGTTGAGAAAGCTTGTG 1584
 QY 1311 --CCACCTGCTGAGGCTGATGATG 1334
 Db 1585 CACCCCTGCTGAGGCTGATGATG 1610

RESULT 8
 AAC76156
 ID AAC76156 standard; cDNA; 2988 BP.
 AC AAC76156;
 XX 08-FEB-2001 (first entry)
 DE Human ORFX ORF111 polynucleotide sequence SEQ ID NO:3421.
 XX
 XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
 KM vulnery; antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 KM anticonvulsant; osteopathic; antirheumatic; immunosuppressant; cardiant;
 KM immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
 KM hypotensive; dermatological; immunosuppressive; antihypertensive;
 KM antiviral; antibacterial; antifungal; antineoplastic; antitumor;
 KM antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
 KM neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KM cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KM cholesterol ester storage; systemic lupus erythematosus; infection;
 KM severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KM allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KM bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KM thrombosis; contraceptive; ss.
 XX
 XX Homo sapiens.
 OS
 XX
 PN WO200058473-A2.

XX 05-OCT-2000.
 PD 31-MAR-2000; 2000WO-US08621.
 XX 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX (CURA-) CURAGEN CORP.
 XX Shinketsu RA, Leach M;
 PI WPI; 2000-602362/57.
 DR P-PSDB: AAB41947.
 DR Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX Claim 5; Page 2597-2599; 5507pp; English.
 XX AAC74446 to AAC77606 encode the proteins given in AAB40337 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytosolic; hepatotropic; vulnery;
 CC antiparasitic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopathic; anticonvulsant; antirheumatic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antineoplastic; antibacterial; antiviral; antitumor; antihypertensive;
 CC antineoplastic; gene therapy; cancer; proliferative disorder; hypertension;
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease, to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX Sequence 2988 BP; 613 A; 916 C; 842 G; 617 T; 0 other.
 SQ
 Query Match 26.8%; Score 547.2; DB 21; Length 2988;
 Best Local Similarity 72.0%; Pred. No. 3.9e-150;
 Matches 730; Conservative 0; Mismatches 278; Indels 6; Gaps 1;
 QY 248 GAGGCGATGACAGCTGCTGCTCTCTTCCGCGGCTGTGGCTCTTCTTGTGGGCGAG 307
 Db 1276 GATGCTGTCACAGCTGCTGCTCTCTTCCGCGGCTGTGGCTCTTCTTGTGGGCGAG 1335
 QY 308 CTTTGGGCGAGCCAGATGTTTCTTACCTGATGAGCCCGGCTGACAGTGTGATG 367
 Db 1336 CTTTGGGCGAGCCAGATGTTTCTTACCTGATGAGCCCGGCTGACAGTGTGATG 1395
 QY 368 ACCTTCAAGCAGACACCGCTGATGCTGACATGCTGTGCGGATGTGATACGGGCC 427
 Db 1396 ACCCTGTCGAGGCGAGCGGCGGCAAGCTGACATGCGCGGCGGACATGATCGCTCT 1455
 QY 428 GTCTTCTTGGCGAGATGAGCGCTTGTGATGAGCCCTGATGAGTGTGCTGCGGAGACAG 487
 Db 1456 ATCTTTTGTGCGAGATGAGCGCTTGTGATGAGCCCTGATGAGTGTGCTGCGGAGACAG 1515
 QY 488 TCCAGCCTCTTCTGATGAGGAGAAACAGCCGGGCTGTGCTGACCTGCTGAGATC 547
 Db 1516 TCCAGCCTCTTCTGATGAGGAGAAACAGCCGGGCTGTGCTGACCTGCTGAGATC 1575
 QY 548 ATCCACAGATGAATATATCCCGGGCTGATGAGGCTCTGTGACAGTCAACAGCCC 607

QY 154 CCTGCTGTTTGTGTTTCCAGATGGCATCTTGCTATTTCTCCACATGTACGCCA 213
 DB 95 CCGTGTGTTTGTGTTTCCAGATGGCATCTTGCTATTTCTCCACATGTACGCCA 154
 QY 214 CAACATGACCTCCCTGTCTATGAGGACAGCCCGGACATGACAGCTGTCTGTG 273
 DB 155 CAACATGACCTCCCTGTCTATGAGGACAGCCCGGACATGACAGCTGTCTGTG 214
 QY 274 TTCTGAGGCTGTGCTCTTTTGTGAGGAGCTTTTGGGACAGCACCAGATGTTT 333
 DB 215 TTCTGAGGCTGTGCTCTTTTGTGAGGAGCTTTTGGGACAGCACCAGATGTTT 274
 QY 334 CTACCTGATGAGACCCGCTGACAGTGTGATGACCTTCAAGACAGACCCGCTGAT 393
 DB 275 CTACCTGATGAGACCCGCTGACAGTGTGATGACCTTCAAGACAGACCCGCTGAT 334
 QY 394 GGTGACATGCTGTGAGGAGATGATACGGCCGCTCTTGTGTGGAATAGGCTCT 453
 DB 335 GGTGACATGCTGTGAGGAGATGATACGGCCGCTCTTGTGTGGAATAGGCTCT 394
 QY 454 TGATGCTACATGAGAACCTGTCTCCCGAGACAGTCCAGCTCTTCACTGGAGAACAG 513
 DB 395 TGATGCTACATGAGAACCTGTCTCCCGAGACAGTCCAGCTCTTCACTGGAGAACAG 454
 QY 514 CCGGCGCCCTGTCTGTGACCTGCTGTGATCATCCACAGATGAATATCCCGCG 573
 DB 455 CCGGCGCCCTGTCTGTGACCTGCTGTGATCATCCACAGATGAATATCCCGCG 514
 QY 574 GGTGACATGCTGTGAGGAGATGATACGGCCGCTCTTGTGTGGAATAGGCTCT 633
 DB 515 GGTGACATGCTGTGAGGAGATGATACGGCCGCTCTTGTGTGGAATAGGCTCT 574
 QY 634 CTCTACAGCAGAGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTACCC 693
 DB 575 CTCTACAGCAGAGTGTGCTCAAGAGAGTGGCTTCTTCAACCTGACAGTCCCTACCC 634
 QY 694 GGTGAGAAAGACCCCTCCCTCAACCTGATATCGTGACCTGTGTCCGGAGCCCGGCG 753
 DB 635 GGTGAGAAAGACCCCTCCCTCAACCTGATATCGTGACCTGTGTCCGGAGCCCGGCG 694
 QY 754 CGTGTCCGTTCCCGAGACGACAAAGAGAGATCATGATGACATGCGATTGTAT 813
 DB 695 CGTGTCCGTTCCCGAGAGACGACAAAGAGAGATCATGATGACATGCGATTGTAT 754
 QY 814 GGGGACGATGAGCAAAAATCTCAAGAGAGAGAGCAACCTACTATGTATGACAGTCA 872
 DB 755 GGGGACGATGAGCAAAAATCTCAAGAGAGAGAGCAACCTACTATGTATGACAGTCA 814
 QY 873 -TCTGCCAAGGACG- TGGAGATCTACAGAC 904
 DB 815 TCTGCCAAGGACGCTTGGAGAACTACAAAGAC 848
 RESULT 7
 AA220793
 ID AA220793 standard; DNA; 1926 BP.
 AC AA220793;
 XX
 DT 08-DEC-1999 (first entry)
 XX
 DE Mouse glycosyl sulfotransferase-3 coding sequence.
 XX
 KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
 KW selectin binding interaction; inflammation; lymphocyte homing; mouse;
 KW secondary lymph organ; ss.
 XX
 OS Mus sp.
 XX
 PN W09949018-A1.
 XX
 PD 30-SEP-1999.
 XX

PF 26-FEB-1999; 99MO-US04316.
 XX
 PR 20-MAR-1998; 98US-0045284.
 PR 12-NOV-1998; 98US-0190911.
 XX
 PA (REGC) UNIV CALIFORNIA.
 PA (SYNT) SYNTAX USA INC.
 XX
 PI Bistrup A, Rosen SD, Tangemann K, Hemmerlich S;
 XX
 DR WPI, 1999-580442/49.
 DR P-PSDB; AAY39919.
 XX
 PT Human and murine glycosyl sulfotransferase 3 and related
 PS polynucleotides
 XX
 PS Claim 4; Fig 3; 59pp; English.
 XX
 CC This sequence encodes the mouse glycosyl sulfotransferase-3 (GST-3) of
 CC the invention. The nucleic acid sequences, probes and primers derived
 CC from these, proteins and antibodies are useful in detecting homologues.
 CC The sequences, antibodies and methods are useful in the diagnosis and
 CC treatment of diseases associated with selectin binding interactions,
 CC including conditions associated with or resulting from the homing of
 CC leukocytes to sites of inflammation and the normal homing of lymphocytes
 CC to secondary lymph organs.
 CC
 SQ Sequence 1926 BP; 426 A; 520 C; 484 G; 496 T; 0 other;
 Query Match. 36.4%; Score 743.2; DB 20; Length 1926;
 Best Local Similarity 76.5%; Pred. No. 4,6e-208;
 Matches 938; Conservative 0; Mismatches 283; Indels 5; Gaps 2;
 QY 111 TCTTCACCTTACAGCAATGCTACTGCTAAATAAAGCTCCGTGTTCTGTT 170
 DB 388 TCTTCCTTCCGAGATGATGCTTTTAAGAAAGAGGCTGCTGATGTTCTGTT 447
 QY 171 CCCAGATGGCATCTTGTGCTATTTCTCCATGTACAGCCACAACATCACTCCGT 230
 DB 448 CCCAGATGATCTGTGAGCTCTCTTCATCATATGCTCCACAGACACTTCCAG 506
 QY 231 CATAGAGGACAGCCCGAGCCAGTCTGTTGTTGTTCTCTGCTGCTGCT 290
 DB 507 --AGGAGGAGTCCAGAGAGCCCGTCTGATGCTGTGCTGCTGCTGCTGAT 564
 QY 291 CTTCCTTGTGGGAGCTTTTGGGACAGCACCAGATGTTTCTACTGATGAGCCG 350
 DB 565 CCTCTTTTGTGGAGACGTTTTCGGGACACCCGAGTGTCTTACCTGATGAGCCG 624
 QY 351 CCTGACAGTGTGATGACCTTCAAGAGACAGCAGCTGATGCTGACATGCTGTC 410
 DB 625 CCTGACATGTGATGATGATCTTCAACAGCAGACAGCCTGAAAGCTGACATGCTGTC 684
 QY 411 GGGATGATGAGGAGGCGGCTCTTGTGAGATGAGGCTCTTGAATGCTACATGAGAC 470
 DB 685 GGGATGATGAGGAGGCGGCTCTTGTGAGATGAGGCTCTTGAATGCTACATGAGAC 744
 QY 471 CTGTTCCCGGAGACAGTCCAGCCTCTTCACTGAGAGAAACGCGGCTGTGTTG 530
 DB 745 CAGGCCCCGGAAGAGTCCAGGCTCTTCAAGAGGAGAAAGCCGCGCTGTGTCAG 804
 QY 531 CACCTGCTGTGATCATCTCCACAAGATGAATATCTCCCGGGCTCACTGACAGCTCC 590
 DB 805 CCGCTGTGTGATCACTTCTTCCGCGCAGAGATGAGCTACCCAAAGACCTCAAGCTCC 864
 QY 591 TGTGAGTCAACAGCCCTTGTGAGTGTGAGAGAGCTGCGGCTCTTCAAGACAGCTG 650
 DB 865 TGTGAGTCAACAGCCCTTGTGAGTGTGAGAGAGCTGCGGCTCTTCAAGACAGCTG 924
 QY 651 TGTCAAGAGAGTGTGCTTCAACCTGACAGTCCCTTACCGGCTGTAAGAACCCCT 710
 DB 925 TACTCAAGAGAGTGTGCTTCAAGCTGACAGCCCTTATCACTACTACAGGACCCCT 984

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX WPI: 2001-524255/58.

XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -

XX Claim 2; SEQ ID NO 263; 1380bp + sequence listing; English.

XX The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is the nucleotide
CC sequence of the 5'-end of a cDNA provided in the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;

Query Match 37.4%; Score 763.6; DB 22; Length 877;

Best Local Similarity 97.7%; Pred. No. 2.8e-214;

Matches 795; Conservative 0; Mismatches 16; Indels 3; Gaps 2;

```

QY 94 AAGCCCAAGCCACAGAGCTTCCACTGACGACAAATGTACTGCTTAAAAAAGAAAGCT 153
DB 35 AAGCCCGCTTGTCAAGGCTTCCACTTCAGCAACAATGCTACTGCTTAAAAAAGAAAGCT 94
QY 154 CCTGCTGTTTCTGTTTCCAGATGAGCATCTTGGCTATCTTCCATGATGACAGCA 213
DB 95 CCTGCTGTTTCTGTTTCCAGATGAGCATCTTGGCTATCTTCCATGATGACAGCA 154
QY 214 CAACATGAGCTCCTGCTATGAAAGCAGACCCGAGCGATGCACTGCTGTTGTC 273
DB 155 CAACATGAGCTCCTGCTATGAAAGCAGACCCGAGCGATGCACTGCTGTTGTC 214
QY 274 TTCTGGGGCTGCTGCTCTTCTTGTGGGGGACAGCTTTGGGAGCCCAAGATGTTT 333
DB 215 TTCTGGGGCTGCTGCTCTTCTTGTGGGGGACAGCTTTGGGAGCCCAAGATGTTT 274
QY 334 CTACCTGATGAGCCCGCTGCGACGCTGTGATGACCTTCAAGAGAGACCGCTGGAT 393
DB 275 CTACCTGATGAGCCCGCTGCGACGCTGTGATGACCTTCAAGAGAGACCGCTGGAT 334
QY 394 GCTGACATGAGCTGCTGGGGATGATGAGGCGCTCTTGTGCGACATGAGCGCTT 453
DB 335 GCTGACATGAGCTGCTGGGGATGATGAGGCGCTCTTGTGCGACATGAGCGCTT 394
QY 454 TGATGCTACATGAGAACCTGCTCCCGAGACATCCAGACCTTTTCACTGAGGAGAG 513
DB 395 TGATGCTACATGAGAACCTGCTCCCGAGACATCCAGACCTTTTCACTGAGGAGAG 454
QY 514 CCGGGCCCTGTGCTGACCTGCTGATCATCCCAAGATGAAATCATCCCCCG 573
DB 455 CCGGGCCCTGTGCTGACCTGCTGATCATCCCAAGATGAAATCATCCCCCG 514
QY 574 GGTCTACTGAGGCTCTGTGCACTCAACAGCCCTTTGAGTGTGAGAAAGGCTGCG 633
DB 515 GGTCTACTGAGGCTCTGTGCACTCAACAGCCCTTTGAGTGTGAGAAAGGCTGCG 574
QY 634 CTCTTACAGCAGCTGCTGCTCAAGAGAGTGCCTTCTTCACTGACAGTCCCTTACCC 693
DB 575 CTCTTACAGCAGCTGCTGCTCAAGAGAGTGCCTTCTTCACTGACAGTCCCTTACCC 634
QY 694 GCTGCTGAAAGAGCCCTCCCAACCTGCAATATGCTGACCTGCTCGGGAGCCCGGGGC 753
DB 635 GCTGCTGAAAGAGCCCTCCCAACCTGCAATATGCTGACCTGCTCGGGAGCCCGGGGC 694

```

```

QY 754 CGTGTCCGTTCCCGAGAACCCACAAAGGAGGATCTCATGTGACACTGCGCATTTGAT 813
DB 695 CGTGTCCGTTCCCGAGAACCCACAAAGGAGGATCTCATGTGACACTGCGCATTTGAT 754
QY 814 GGGCAGCATGAGCAAAAACCTCAGAGAGAGACCAACCTTACTATGTGATGAGGTGA- 872
DB 755 GGGCAGCATGAGCAAAAACCTCAGAGAGAGACCAACCTTACTATGTGATGAGGTGA 814
QY 873 -TCTGCCAAGCCAGC-TGAGATCTACAGACC 904
DB 815 TCTGCCAAGCCAGCCTTGAGAGACTACAGAAC 848

```

RESULT 6
AAK93921

ID AAK93921 standard; cDNA; 877 BP.

AC AAK93921;

XX 06-NOV-2001 (first entry)

XX Human cDNA clone representative sequence; SEQ ID NO: 2381.

XX Human; full length cDNA; cDNA synthesis; oligo-capping; ss.

XX Homo sapiens.

XX EP1130094-A2.

XX 05-SEP-2001.

XX 07-JUL-2000; 2000EP-0114089.

XX 08-JUL-1999; 99JP-0194486.

XX 11-JAN-2000; 2000JP-0118774.

XX 02-MAY-2000; 2000JP-0185765.

XX (HELI-) HELIX RES INST.

PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;

XX WPI: 2001-524255/58.

DR 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -

PS Example 11; SEQ ID NO 2381; 1380bp + sequence listing; English.

CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence was used as the
CC representative sequence from a human clone which was used in
CC homology searches to identify the clone.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.

XX Sequence 877 BP; 183 A; 266 C; 221 G; 203 T; 4 other;

Query Match 37.4%; Score 763.6; DB 22; Length 877;

Best Local Similarity 97.7%; Pred. No. 2.8e-214;

Matches 795; Conservative 0; Mismatches 16; Indels 3; Gaps 2;

```

QY 94 AAGCCCAAGCCACAGAGCTTCCACTGACGACAAATGTACTGCTTAAAAAAGAAAGCT 153
DB 35 AAGCCCGCTTGTCAAGGCTTCCACTTCAGCAACAATGCTACTGCTTAAAAAAGAAAGCT 94

```

PT Intestinal G1cMac 6-sulfotransferase -
XX
PS Claim 19; Fig 4; 98pp; English.

CC The present invention provides a method of modifying an acceptor molecule
CC by contacting the acceptor with an isolated
CC betal, 3-N-acetylglucosaminyltransferase (betal, 3gnt) or an active
CC fragment, where betal, 3gnt directs expression of a MECA-79 antigen. The
CC invention also provides a method of treating or preventing an
CC L-selectin-mediated condition by reducing the expression or activity of a
CC betal, 3gnt that directs expression of a MECA-79 antigen. This can be done
CC by administering to the subject an oligosaccharide L-selectin antagonist
CC that inhibits binding of L-selectin to a MECA-79 antigen, for example by
CC administering antibody material that specifically binds betal, 3gnt,
CC and/or a betal, 3gnt antisense nucleic acid molecule. L-selectin
CC sulfotransferase-2 (LSSST-2) also directs MECA-79 antigen expression.
CC Alternatively, the expression or activity of LSSST-2 or its active
CC fragment can be reduced in combination with reducing the expression or
CC activity of betal, 3gnt. The method is useful for treating L-selectin
CC mediated conditions such as Crohn's disease and ulcerative colitis,
CC inflammatory disorders of the skin such as allergic contact dermatitis
CC psoriasis and Lichen planus, lymphomas, chronic pneumonia, delayed-type
CC hypersensitivity reactions, diabetes and hyperplastic thymus. This
CC sequence represents cDNA encoding human LSSST-2.

Sequence 1333 BP; 286 A; 393 C; 336 G; 318 T; 0 other;

Query Match	61.88;	Score 1262;	DB 24;	Length 1333;
-------------	--------	-------------	--------	--------------

```
Matches 1262; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY	107	AAAGTCTTCCACTGACGACATAGCTACTGCTAATAAATAAGAGCTCTGCTGTTCTG	168
Db	72	AAAGTCTTCCACTGACGACAAATGCTACTGCTAAAAAATAAGAGCTCTGCTGTTCTG	131
QY	167	GTTTCCCAAGATGAGCCATCTTGCTATCTTCTTCCACATGTACACCAACATCAGCTCC	226
Db	132	GTTTCCCAAGATGAGCCATCTTGCTATCTTCTTCCACATGTACACCAACATCAGCTCC	191
QY	227	CTGTCTATGAAGGACACAGCCGACGATGACAGTGTGCTTCTGTTCTCTGCGCTCT	286
Db	192	CTGTCTATGAAGGACACAGCCGACGACGATGACAGTGTGCTTCTGTTCTCTGCGCTCT	251
QY	287	GGCTCTTCTTTTGTGGGGCAGCTTTTGTGGGACGACCCAGATGTTTCTACTATGAG	346
Db	252	GGCTCTTCTTTTGTGGGGCAGCTTTTGTGGGACGACCCAGATGTTTCTACTATGAG	311
QY	347	CCGCGCTGGACAGTGTGATGACTTTCACAGACAGACCCGCTGATGTGACATGGCT	406
Db	312	CCGCGCTGGACAGTGTGATGACTTTCACAGACAGACCCGCTGATGTGACATGGCT	371
QY	407	GTCGGGAATCTATACGGGCGCTCTTCTGTGTCGACATGAGCGTCTTGATGCGCTACATG	466
Db	372	GTCGGGAATCTATACGGGCGCTCTTCTGTGTCGACATGAGCGTCTTGATGCGCTACATG	431
QY	467	GAACCTGTGTCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAAACAGCCGGGCCCTGTGT	526
Db	432	GAACCTGTGTCCCGGAGACAGTCCAGCCTCTTTCAGTGGGAAACAGCCGGGCCCTGTGT	491
QY	527	TCGACACCTGCGTGTACATCATCCACAAGATGAATCATCCCGGGGCTCACTGTAGG	586
Db	492	TCGACACCTGCGTGTACATCATCCACAAGATGAATCATCCCGGGGCTCACTGTAGG	551
QY	587	CTCCTGTGACATCACAAGCCCTTTGAGGTGTGGAAAGGCTCGCCGCTCTCTACAGCCAC	646
Db	552	CTCCTGTGACATCACAAGCCCTTTGAGGTGTGGAAAGGCTCGCCGCTCTCTACAGCCAC	611
QY	647	GTCGTCTCAAGAGAGGTGGCTTCTTTCACACCGACGATCCCTCTACCCGCTGTAAAGAC	706
Db	612	GTCGTCTCAAGAGAGGTGGCTTCTTTCACACCGACGATCCCTCTACCCGCTGTAAAGAC	671
QY	707	CCCTCCCTAACCTGCATATGTGCACCTGTGCGGGAGCCCGGGGCGTGTGTCGTTCC	766

Db	672	CCGCTCCCTCAACCTGCATATATGTGCACCTGGTCCGGGAGCCCCCGGGCCGTGTTCCTGTTCC	731
Oy	767	CGAAGAGCGCAACAAAGGAGATCTCATATGTTACAGTGTGCATTTGATGAGGGCAGCATAG	826
Db	732	CGAAGAGCGCAACAAAGGAGATCTCATATGTTACAGTGTGCATTTGATGAGGGCAGCATAG	791
Oy	827	CAAAAGCTCAAGAGAGAGAGACCAACCCCTACTATGTGATGCAGGTCATCTGCCAAAGCCAG	886
Db	792	CAAAAGCTCAAGAGAGAGAGACCAACCCCTACTATGTGATGCAGGTCATCTGCCAAAGCCAG	851
Oy	887	CTGAGAGATCTACAAAGACCATCCAGTCTCTTGCCCAAGGCCCTGCAGAAAGCTTACCTGCTT	946
Db	852	CTGAGAGATCTACAAAGACCATCCAGTCTCTTGCCCAAGGCCCTGCAGAAAGCTTACCTGCTT	911
Oy	947	GTGGGCGATATGAGAGACCTGGCTCGAGCCCCCTGGGCCCAAGACTTCCCGAATGTATGAATTC	1006
Db	912	GTGGGCGATATGAGAGACCTGGCTCGAGCCCCCTGGGCCCAAGACTTCCCGAATGTATGAATTC	971
Oy	1007	GTGGGATTTGGAATTCCTTGCCCATCTTCAGACCTGGGTGCATTAACATCACCCGAGGCAAG	1066
Db	972	GTGGGATTTGGAATTCCTTGCCCATCTTCAGACCTGGGTGCATTAACATCACCCGAGGCAAG	1031
Oy	1067	GGCATGGGTGACCAACGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCT	1126
Db	1032	GGCATGGGTGACCAACGCTTTCCACACAAATGCCAGGATGCCCTTAATGTCTCCAGGCT	1091
Oy	1127	TGGGGCTGGGTGTTTGGCCCATGAAAGGTTTCTGACCTTCAGAAAGCCTGGGGATGCC	1186
Db	1092	TGGGGCTGGGTGTTTGGCCCATGAAAGGTTTCTGACCTTCAGAAAGCCTGGGGATGCC	1151
Oy	1187	ATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAACAAGAGAGAAACCTGTGGCTG	1246
Db	1152	ATGAATTTGCTGGGCTACCGCCACGTCAGATCTGAACAAGAGAGAAACCTGTGGCTG	1211
Oy	1247	GATCTTCTGTCTACTGTGACCTGTCCCTGAGCAATCCACTAAGAGGTTGGAAGGCTTT	1306
Db	1212	GATCTTCTGTCTACTGTGACCTGTCCCTGAGCAATCCACTAAGAGGTTGGAAGGCTTT	1271
Oy	1307	GCTGCGCACCTGGTGTACAGCCCTCAGTCACTTCTCTGAAATGCTTGTGAGCCTTGGCTACAT	1366
Db	1272	GCTGCGCACCTGGTGTACAGCCCTCAGTCACTTCTCTGAAATGCTTGTGAGCCTTGGCTACAT	1331
Oy	1367	CT 1368	
Db	1332	CT 1333	
RESULT 5			
AAK91803			
ID	AAK91803 standard; cDNA: 877 BP.		
XX	AAK91803:		
XX	AC		
XX	AAK91803: (first entry)		
XX	06-NOV-2001		
DE	Human cDNA 5'-end sequence, SEQ ID NO: 263.		
XX	Human; full length cDNA; cDNA synthesis; oligo-capping; ss.		
XX	Homo sapiens.		
XX	OS		
XX	EP1130094-A2.		
PD	05-SEP-2001.		
XX	07-JUL-2000; 2000EP-0114089.		
XX	08-JUL-1999; 99JP-0194486.		
PR	11-JAN-2000; 2000JP-0118774.		
PR	02-MAY-2000; 2000JP-0183765.		
XX	(HELI-) HELIX RES INST.		
XX			

TTCC	731
TTGAG	826
TTTT	TTTT
TTGAG	791
CCAC	886
TTTT	TTTT
CCAC	851
GGCTT	946
TTTT	TTTT
GGCTT	911
ATTC	1006
TTTT	TTTT
ATTC	971
GGGT	1091
TTTT	TTTT
GGGT	1126
CAAG	1031
TTTT	TTTT
CAAG	1066
TTTT	TTTT
TTGCC	1186
TTGCC	1151
GGCTG	1246
TTTT	TTTT
GGCTG	1211
CTTT	1306
TTTT	TTTT
CTTT	1271
ACAT	1366
TTTT	TTTT
ACAT	1331

Db 508 TGAATCCCTACATGGAACCTGGTCCCGGAGACAGTCCAGCTCTTTTCAGTGGGAGAACAG 567
 QY 514 CCGGGCCCTGTGTCTGACCTGGCTGTGACATCATCCACAGATGAATCATCCCGG 573
 Db 568 CCGGGCCCTGTGTCTGACCTGGCTGTGACATCATCCACAGATGAATCATCCCGG 627
 QY 574 GGTCTACATGAGGCTCTGTGACATCAACAGCCCTTTGAGGTGTGAGAAAGCCCTGGC 633
 Db 628 GGTCTACATGAGGCTCTGTGACATCAACAGCCCTTTGAGGTGTGAGAAAGCCCTGGC 687
 QY 634 CTCTACAGCAGCTGGTGTGCAAGAGGTGGCTCTTCACTCACTGCACTCCCTCAACC 693
 Db 688 CTCTACAGCAGCTGGTGTGCAAGAGGTGGCTCTTCACTCACTGCACTCCCTCAACC 747
 QY 694 GCTGCTGAAAGACCCCTCCCTCAACCTGATATGTCGACCTGGTCCGGGACCCCGGGC 753
 Db 748 GCTGCTGAAAGACCCCTCCCTCAACCTGATATGTCGACCTGGTCCGGGACCCCGGGC 807
 QY 754 CTTGTTCCTGTCCTCCGAGACGACCAAGAGGATCTCATGATTTGACAGTGCATTTGAT 813
 Db 808 CTTGTTCCTGTCCTCCGAGACGACCAAGAGGATCTCATGATTTGACAGTGCATTTGAT 867
 QY 814 GGGGACATGAGGCAAAACTCAAGAGAGAGACCAACCTACTATGATGATGAGGTGAT 873
 Db 868 GGGGACATGAGGCAAAACTCAAGAGAGAGACCAACCTACTATGATGATGAGGTGAT 927
 QY 874 CTGCAAGACCAAGCTGAGATCTCAAGAACCATTCAGTCTTGGCCCAAGGCCCTGCAAGA 933
 Db 928 CTGCAAGACCAAGCTGAGATCTCAAGAACCATTCAGTCTTGGCCCAAGGCCCTGCAAGA 987
 QY 934 ACCTTACCTGCTTGTGGCTATGAGACCTGCTGAGACCCCTGTGGCCCAAGCTTCCCG 993
 Db 988 ACCTTACCTGCTTGTGGCTATGAGACCTGCTGAGACCCCTGTGGCCCAAGCTTCCCG 1047
 QY 994 AATGTATGAATTCGTGGGATTTGGAATTTGCCCCATCTTCAAGCTGGGTGATATCAT 1053
 Db 1048 AATGTATGAATTCGTGGGATTTGGAATTTGCCCCATCTTCAAGCTGGGTGATATCAT 1107
 QY 1054 CACCCGAGGCAAGGGCATGGGTGACACAGCTTTCCACACAATCCAGGGATGCCCTTAA 1113
 Db 1108 CACCCGAGGCAAGGGCATGGGTGACACAGCTTTCCACACAATCCAGGGATGCCCTTAA 1167
 QY 1114 TGTCTCCAGGCTTTGGGCTGTGTGCTTGGCTATGAAAAGTTTCTGACTTCAGAAAGC 1173
 Db 1168 TGTCTCCAGGCTTTGGGCTGTGTGCTTGGCTATGAAAAGTTTCTGACTTCAGAAAGC 1227
 QY 1174 CTGTGCGAGTGCATGATTTGTGCGGCTACCGGCTACGTCGATCTGAAACAAGAACAG 1233
 Db 1228 CTGTGCGAGTGCATGATTTGTGCGGCTACCGGCTACGTCGATCTGAAACAAGAACAG 1287
 QY 1234 AATCCTTGTGCTGATTTCTGTCTACCTGATGCTGCTGAGCAAAATCCACTAAGAGG 1293
 Db 1288 AATCCTTGTGCTGATTTCTGTCTACCTGATGCTGCTGAGCAAAATCCACTAAGAGG 1347
 QY 1294 TTGGAAGGCTTTGCTGCAACCTGCTGACCCCTCACTCACTTCTCTGATGCTTCTGA 1353
 Db 1348 TTGGAAGGCTTTGCTGCAACCTGCTGACCCCTCACTCACTTCTCTGATGCTTCTGA 1407
 QY 1354 GCTTGGCTACATCTGAGCCTTAATACATGCTGTTGGTATCACACTGAGTGTGAGT 1413
 Db 1408 GCTTGGCTACATCTGAGCCTTAATACATGCTGTTGGTATCACACTGAGTGTGAGT 1467
 QY 1414 TGTGTCCACAGTGTCTCAAGCAGAGACTTTTGTGTCTGATGCTGTGTGTAGAAACAG 1473
 Db 1468 TGTGTCCACAGTGTCTCAAGCAGAGACTTTTGTGTCTGATGCTGTGTGTGTAGAAACAG 1527
 QY 1474 ACTGGGAACTTATGTGAGACAGACATCCACAGTGAACAGAGGATTTGCTTCTTC 1533
 Db 1528 ACTGGGAACTTATGTGAGACAGACATCCACAGTGAACAGAGGATTTGCTTCTTC 1587
 QY 1534 TTTTCTTGAATCTTCTGCTGCTGGGAGACTTCAGAGACTTTTGGCCCTGGAGGCTATTTAA 1593
 Db 1588 TTTTCTTGAATCTTCTGCTGCTGGGAGACTTCAGAGACTTTTGGCCCTGGAGGCTATTTAA 1647

QY 1594 GCACGACACAGTATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTGCCCAATG 1653
 Db 1648 GCACGACACAGTATCAGTGAATGATTCATTAACCTCCCTGTCACATCTTGCCCAATG 1707
 QY 1654 GGGAAATGATTTTCCACCAAGAGCTCACAGCATTTTTCACAGAGATGCAATTTGAG 1713
 Db 1708 GGGAAATGATTTTCCACCAAGAGCTCACAGCATTTTTCACAGAGATGCAATTTGAG 1767
 QY 1714 CCGTTGGATGCCCAATGGGATTCAGAGAGAGTGGGAACAAGTTGATGCTACTT 1773
 Db 1768 CCGTTGGATGCCCAATGGGATTCAGAGAGAGTGGGAACAAGTTGATGCTACTT 1826
 QY 1774 ATGACCTTGACCAT-ACAGTATGCGTATGCAAGAAATTTGAACAAATCTCTG-ACAA 1831
 Db 1827 ATGACCTTGACCATGACATGACATGATGATATGAAATTTGAACAAATCTCTGACAAA 1886
 QY 1832 AGAGCAACCTTTAAGTTACAGAGTGGCTGGCTTGTATTTGAATATCA-TTTCCTTTG 1890
 Db 1887 AGAGCAACCTTTAAGTTACAGAGTGGCTGGCTGCAATTTGAATATCACTTCCCTCTG 1946
 QY 1891 CATTTTCCATTAATGAAACCTTTGACCTGTGAACTTGCAATCTGTAATATAA 1950
 Db 1947 CATTTTCCATTAATGAAACCTTTGACCTGTGAACTTGCAATCTGTAATATAA 2005
 QY 1951 TTCCCAATTAAG 1962
 Db 2006 TTCCCAATTAAG 2017

RESULT 4

AAS16947
 AAS16947 standard; cDNA; 1333 BP.

AAS16947;
 12-MAR-2002 (first entry)

Human L-selectin sulfotransferase-2 (LSST-2) cDNA.

Human: beta1,3gal: beta1,3-N-acetylgalactosaminyltransferase; MECA-79; ss;
 L-selectin; L-selectin sulfotransferase-2; Crohn's disease; diabetes;
 ulcerative colitis; inflammatory skin disorder; psoriasis; lichen planus;
 allergic contact dermatitis; lymphoma; chronic pneumonia; LSST-2;
 delayed-type hypersensitivity reaction; hyperplastic thymus; antidiabetic;
 antiinflammatory; antipsoriatic; antidiabetic; dermatological;
 antiallergic.

Homo sapiens.

Key Location/Qualifiers
 CDS 111..1253
 FT /*tag= a
 FT /product= "Human LSST-2"

WO200185177-A1.

15-NOV-2001.

10-MAY-2001; 2001WO-US15452.

11-MAY-2000; 2000US-0569320.

(BURN-) BURNHAM INST.

Fukuda M, Yeh J, Hirooka N;

WPI; 2002-075226/10.

P-PSDB; AAU11274.

New enzyme, useful for modifying acceptor molecule, comprises an
 isolated L-selectin sulfotransferase-2 that directs expression of
 L-selectin ligand antigen, MECA-79 in Chinese hamster ovary cells, or

```

Db 1099 TCACCCGAGGCAAGGGCATGGGTGACACGCTTTCCACAAATGCGAGGATGCCCTTA 1158
QY 1113 ATGTCTCCAGGCTTGGCGCTGGTCTTTGCTCCATGAAAGTTTCTGCATTCAGAAAG 1172
Db 1159 ATGTCTCCAGGCTTGGCGCTGGTCTTTGCTCCATGAAAGTTTCTGCATTCAGAAAG 1218
QY 1173 CCTGTGGCGATGCGATGAATTTGCTGGGTACCGCGACGTCAAGATGGAACAAGACAGA 1232
Db 1219 CCTGTGGCGATGCGAAGATTTGCTGGGTACCGCGACGTCAAGATGGAACAAGACAGA 1278
QY 1233 GAAACCTGTGGTGGATCTTGTCTACTGACGTGCTCCCTGTGAGAAATCCCATAGAGG 1292
Db 1279 GAAACCTGTGGTGGATCTTGTCTACTGACGTGCTCCCTGTGAGAAATCCCATAGAGG 1338
QY 1293 GTTGAGAGGCTTTGGTGGACCTGGTGTGACGCTTGTGACCTTCTCTGAATGCTTCTG 1352
Db 1339 GTTGAGAGGCTTTGGTGGACCTGGTGTGACGCTTGTGACCTTCTCTGAATGCTTCTG 1398
QY 1353 AGCCTTGCTTACATCTCTGAGCCTTAACATGATGTCTGTGGGTATCACACTGAGTGTAG 1412
Db 1399 AGCCTTGCTTACATCTCTGAGCCTTAACATGATGTCTGTGGGTATCACACTGAGTGTAG 1458
QY 1413 TTGTGTCCACAGCTGCTCAAGCAGAGGACTTTGTGTCCATGCTTGTGTAGAAACAA 1472
Db 1459 TTGTGTCCACAGCTGCTCAAGCAGAGGACTTTGTGTCCATGCTTGTGTAGAAACAA 1518
QY 1473 GACTGGGGAACCTTATGTAGCAGACATCCACACAGTAAACAGGATTTGCTCTTCTT 1532
Db 1519 GACTGGGGAACCTTATGTAGCAGACATCCACACAGTAAACAGGATTTGCTCTTCTT 1578
QY 1533 CTTTCTTATCTCTCTGCTGGGCACTTCAAGAACTTTGTGGCTTGAGAGGCTTATTA 1592
Db 1579 CTTTCTTATCTCTCTGCTGGGCACTTCAAGAACTTTGTGGCTTGAGAGGCTTATTA 1638
QY 1593 AGCAGCAGACAGTATCAGTGAATGATCATTAACCTCCCTGTCCACATCTTGCCTCAAT 1652
Db 1639 AGCAGCAGACAGTATCAGTGAATGATCATTAACCTCCCTGTCCACATCTTGCCTCAAT 1698
QY 1653 GGGGAATGATCTTTACCAAAAGAGCTCACACAGATTTTCCACAGATGCGAATTCGA 1712
Db 1699 GGGGAATGATCTTTACCAAAAGAGCTCACACAGATTTTCCACAGATGCGAATTCGA 1758
QY 1713 GCCCTTGAGTCCCAATGGGATTCAGAAAGAGGAGTGGGAACAAGTTGATGCTTACT 1772
Db 1759 GCCCTTGAGTCCCAATGGGATTCAGAAAGAGGAGTGGGAACAAGTTGATGCTTACT 1817
QY 1773 TATGACCTTGACCAT-ACAGCTATCCGTATCAGAAATATGAAACAAATCTCTG-ACAA 1830
Db 1818 TATGACCTTGACCATCAGAGCTATCGTATATGAAATATGAAACAAATCTCTGACAA 1877
QY 1831 AAGAGCAGCTCTTAAGTTCAAGAGTGGCTGGGCTTGAATTTGAATATCA-TTCCCTTT 1889
Db 1878 AAGAGCAGCTCTTAAGTTCAAGAGTGGCTGGGCTTGAATTTGAATATCA-TTCCCTTT 1937
QY 1890 GAATTTCCCATTAAGTAAAGAACTTGAACCTGTGAACCTTGCCATCTGTATATATAA 1949
Db 1938 GAATTTCCCATTAAGTAAAGAACTTGAACCTGTGAAG-TCGCATCTGTATATATAA 1996
QY 1950 ATTCCCAATTAAGTTCTGTTAGATGTCCTTTTATGCTTCTTAATTAATAGAGTA 2009
Db 1997 ATTCCCAATTAAGTTCTGTTAGATGTCCTTTTATGCTTCTTAATTAATAGAGTA 2056
QY 2010 AATGTTCAAT 2018
Db 2057 AATGTTCAAT 2065

```

RESULT 3
 AA220792
 ID AA220792 standard; DNA; 2032 BP.
 XX
 AC AA220792;
 XX

```

DT 08-DEC-1999 (first entry)
XX
DE Human glycosyl sulfotransferase-3 coding sequence.
XX
KW Glycosyl sulfotransferase; GST-3; detection; diagnosis; leukocyte homing;
XX selectin binding interaction; Inflammation; lymphocyte homing; human;
XX secondary lymph organ; ss.
XX
OS Homo sapiens.
XX
PN M09949018-A1.
XX
PD 30-SEP-1999.
XX
PF 26-FEB-1999; 99WO-US04316.
XX
PR 20-MAR-1998; 98US-0045284.
XX
PR 12-NOV-1998; 98US-0190911.
XX
PA (REGC ) UNIV CALIFORNIA.
XX (SYNT ) SYNTAX USA INC.
XX
PI Bistrup A, Rosen SD, Tangemann K, Hemmerlich S;
XX
XX WPI, 1999-580442/49.
XX
XX P-PDSB; AAY39918.
XX
XX Human and murine glycosyl sulfotransferase 3 and related
XX polynucleotides
XX
XX Claim 4; Fig 1; 59pp; English.
XX
PS
CC This sequence encodes the human glycosyl sulfotransferase-3 (GST-3) of
CC the invention. The nucleic acid sequences, probes and primers derived
CC from these, proteins and antibodies are useful in detecting homologues.
CC The sequences, antibodies and methods are useful in the diagnosis and
CC treatment of diseases associated with selectin binding interactions,
CC including conditions associated with or resulting from the homing of
CC leukocytes to sites of inflammation and the normal homing of lymphocytes
CC to secondary lymph organs.
CC
SQ Sequence 2032 BP; 468 A; 569 C; 490 G; 505 T; 0 other;

Query Match 87.4%; Score 1784.8; DB 20; Length 2032;
Best Local Similarity 98.8%; Pred. No. 0;
Matches 1850; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

QY 94 AAGCCCAAGCACAAGGCTCTCCACTTCAGACAAATGCTACTGCTTAAATAATGAAGCT 153
Db 148 AAGCCCGCTCTGCAAGGCTCTCCACTTCAGACAAATGCTACTGCTTAAATAATGAAGCT 207
QY 154 CCGTCTGTTTCTGGTTTCCAGATGGCAATCTTGCTTATTTCCAGATGTACAGCA 213
Db 208 CCGTCTGTTTCTGGTTTCCAGATGGCAATCTTGCTTATTTCCAGATGTACAGCA 267
QY 214 CAACATCAGCTCCCTCTTATGAAGCACAAGCCGAGCGCATGACGTGCTGTCTG 273
Db 268 CAACATCAGCTCCCTCTTATGAAGCACAAGCCGAGCGCATGACGTGCTGTCTG 327
QY 274 TTCCGTGGCCTCGGCTCTTCTTGTGGGGCAGCTTTTGGGACACCCAGATTTT 333
Db 328 TTCCGTGGCCTCGGCTCTTCTTGTGGGGCAGCTTTTGGGACACCCAGATTTT 387
QY 334 CTACCTGATGAGCCCGCTGCGACGTGTGATGACCTTCAAGCAGACAGCCGCTGAT 393
Db 388 CTACCTGATGAGCCCGCTGCGACGTGTGATGACCTTCAAGCAGACAGCCGCTGAT 447
QY 394 GCTGCAATGCTGTGCGGAGATGTATACGGGCGCTCTTCTGTGCGACATGAGCTCTT 453
Db 448 GCTGCAATGCTGTGCGGAGATGTATACGGGCGCTCTTCTGTGCGACATGAGCTCTT 507
QY 454 TGATGCTACATGGAACCTGTGCTCCCGGAGACAGTCCACCTTTCATGTGGAGAACG 513

```

DB 1834 CATTTCCTCCATACATAGACATTTGACCTGTGAAGC-TGCCATCTGTTAATACTATAA 1892
QY 1951 TTCCCAATAAGTCTCTTTAGAAATGCCCTTTTATGCTCTTATATTATAGCAGTAA 2010
DB 1893 TTCCCAATAAGTCTCTTTAGAAATGCCCTTTTATGCTCTTATATTATAGCAGTAA 1952
QY 2011 ATGTTCATTTTATGAGATCTTAA 2034
DB 1953 ATGTTCATTTTATGAGATCTTAA 1976
RESULT 2
AA294211 ID AA294211 standard; cDNA; 2065 BP.
AA294211 AC AA294211;
XX 19-JUN-2000 (first entry)
DE Human transferase TRNSFS-11 cDNA clone 2617407CBI.
XX
XX Transferase; TRNSFS-11; human; antitumor; cell proliferation;
KW inflammation; gastrointestinal disorder; developmental disorder;
KW genetic disorder; neurological disorder; reproductive disorder;
KW smooth muscle disorder; immunological disorder; gene therapy;
diagnosis; N-acetylglucosamine 6-O-sulfotransferase; ss.
XX
XX Homo sapiens.
OS
XX
XX Key Location/Qualifiers
FT CDS 174..1334
FT /tag- a
XX MO200014251-A2.
XX 16-MAR-2000.
XX
XX 09-SEP-1999; 99WO-US20989.
XX
XX 10-SEP-1998; 98US-0150657.
PR 04-NOV-1998; 98US-0186779.
PR 11-MAY-1999; 99US-0133642.
XX
XX (INCY-) INCYTE PHARM INC.
PI Tang YT, Corley NC, Guegler KJ, Baughn MR, Lal P, Yue H;
PI Hillman JL, Azimzal Y;
XX WPI: 2000-256996/22.
DR P-PSDB; AAY79219.
XX
XX Human transferase proteins useful for preventing, diagnosing and
PT treating cancers and developmental, gastrointestinal, genetic,
PT immunological, neurological, reproductive and smooth muscle disorders.
XX
XX Claim 9; Page 104-105; 113pp; English.
XX
XX The present sequence is that of cDNA clone 2617407CBI encoding
CC human transferase TRNSFS-11 (see AAY79219). 1 of 15 claimed human
CC transferase proteins of the invention (see AAY79209-23). The clone
CC was isolated from gall bladder cDNA library GBLANOT01. TRNSFS-11
CC is expressed in dermatologic and gastrointestinal tissues,
CC especially those associated with inflammation and cell
CC proliferation. It shows homology to mouse N-acetylglucosamine
CC 6-O-sulfotransferase. A polynucleotide comprising nucleotides
CC 264-333 or 1272-1331 of the present sequence can be used as a DNA
CC probe. The new human transferases and polynucleotides can be used
CC in the diagnosis, prevention and treatment (including gene therapy
CC and antineoplastic therapy) of cancer, developmental disorders,
CC gastrointestinal disorders, genetic disorders, immunological
CC disorders, neurological disorders, reproductive disorders, and
CC smooth muscle disorders.

XX SQ Sequence 2065 BP; 466 A; 575 C; 491 G; 533 T; 0 other;
Query Match 88.7%; Score 1813; DB 21; Length 2065;
Best Local Similarity 98.6%; Pred. No. 0;
Matches 1902; Conservative 0; Mismatches 20; Indels 7; Gaps 7;
QY 94 AAGCCCAAGCAGCAAGTCTTCCACTTCAGCAATAGCTAGTGCCTAATAAATAAGAACT 153
DB 140 AAGCCGCTTGAAGGCTTCCACTTCAGCAATAGTACTGCCTAATAAATAAGAACT 199
QY 154 CCTGCTTTTCTGTTTCCAGATGCCATCTTGCTCTATTTCTCCATGATGAGCCA 213
DB 200 CCTGCTTTTCTGTTTCCAGATGCCATCTTGCTCTATTTCTCCATGATGAGCCA 259
QY 214 CAACATCAGTCCCTGCTATGAAAGGACAGCCGAGGAGGACGACGCTGCTCTGTC 273
DB 260 CAACATCAGTCCCTGCTATGAAAGGACAGCCGAGGAGGACGACGCTGCTCTGTC 319
QY 274 TTCTGCGCTCTGCTCTCTCTTTGTTGCGGAGCTTTTGGGAGCAGCCAGATCTTT 333
DB 320 TTCTGCGGCTCTGCTCTCTCTTTTGTGGGAGCTTTTGGGAGCAGCCAGATCTTT 379
QY 334 CTACCTGATGAGAGCCGCTGCGACGTTGTGATGACCTTTCAAGCAGAGCAGCCCTGAT 393
DB 380 CTACCTGATGAGAGCCGCTGCGACGTTGTGATGACCTTTCAAGCAGAGCAGCCCTGAT 439
QY 394 GCTGCACATGAGCTGTCGGGAGATGATGAGGCGCTCTTCTGTCGACATGAGCTT 453
DB 440 GCTGCACATGAGCTGTCGGGAGATGATGAGGCGCTCTTCTGTCGACATGAGCTT 499
QY 454 TGATGCCATGATGAACTGCTGCTCCCGAGAGAGTCACGCTCTTCTAGTGGAGAACAG 513
DB 500 TGATGCCATGATGAACTGCTGCTCCCGAGAGAGTCACGCTCTTCTAGTGGAGAACAG 559
QY 514 CCGGCGCTGTTCTGACCTGCTGATGATGATGATGATGATGATGATGATGATGATG 573
DB 560 CCGGCGCTGTTCTGACCTGCTGATGATGATGATGATGATGATGATGATGATGATG 618
QY 574 GGTGCACGAGGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 632
DB 619 GGTGCACGAGGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 678
QY 633 GCTCCTACAGCAGTGTGCTCAAGAGTGTGCTTCAACTGACGCTGCTGCTGCTGCTG 692
DB 679 GCTCCTACAGCAGTGTGCTCAAGAGTGTGCTTCAACTGACGCTGCTGCTGCTGCTG 738
QY 693 CGCTGCTGAAAGACCCCTGCTCAACTGATGATGATGATGATGATGATGATGATGATG 752
DB 739 CGCTGCTGAAAGACCCCTGCTCAACTGATGATGATGATGATGATGATGATGATGATG 798
QY 753 CCGTGTCCGTTCCGAGAACGACCAAGGAGATGATGATGATGATGATGATGATGATGATG 812
DB 799 CCGTGTCCGTTCCGAGAACGACCAAGGAGATGATGATGATGATGATGATGATGATGATG 858
QY 813 TGGGGCAGCATGAGCAAAACTCAAGAGAGGAGCAACCTGATGATGATGATGATGATG 872
DB 859 TGGGGCAGCATGAGCAAAACTCAAGAGAGGAGCAACCTGATGATGATGATGATGATG 918
QY 873 TCTGCAAGAGCAGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 932
DB 919 TCTGCAAGAGCAGCTGAGATGATGATGATGATGATGATGATGATGATGATGATGATG 978
QY 933 AAGCTTACGCTGCTGCTGATGAGGACGCTGCTGAGCCGCTGCTGCTGCTGCTGCTG 992
DB 979 AAGCTTACGCTGCTGCTGATGAGGACGCTGCTGAGCCGCTGCTGCTGCTGCTGCTG 1038
QY 993 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1052
DB 1039 GAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 1098
QY 1053 TCACCCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 1112

XX 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
PS
XX
PS Claim 8; SEQ ID NO 2816; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5' - and 3' - ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a full length
CC human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 1979 BP; 451 A; 542 C; 474 G; 512 T; 0 other;

Query Match 90.9%; Score 1856.8; DB 22; Length 1979;
Best Local Similarity 98.9%; Pred. No. 0;
Matches 1992; Conservative 0; Mismatches 17; Indels 5; Gaps 5;

QY 94 AAGCCCAAGCCACAGGCTTCCACTTCAAGACATGCTACTGCTTAAAAAATGAAGCT 153
DB 35 AAGCCCGCTGTGCAAGGCTCTCCACTTCAAGACATGCTACTGCTTAAAAAATGAAGCT 94
QY 154 CCGTGTGTTGCTGCTTCCAGATGAGCCATCTGCTGCTATCTTCCACATGACAGCA 213
DB 95 CCGTGTGTTGCTGCTTCCAGATGAGCCATCTGCTGCTATCTTCCACATGACAGCA 154
QY 214 CAACATCAGCTCCCTCTATGAAAGCAGACGCCAGCCGATCAGCTGCTGTTGTC 273
DB 155 CAACATCAGCTCCCTCTATGAAAGCAGACGCCAGCCGATCAGCTGCTGTTGTC 214
QY 274 TTCTGTGGGCTCTGGCTCTTCTTTTGTGGGCAAGCTTTTGGCAGACCCAGATGTTT 333
DB 215 TTCTGTGGGCTCTGGCTCTTCTTTTGTGGGCAAGCTTTTGGCAGACCCAGATGTTT 274
QY 334 CTACATGATGAGACCGCCGCTGAGCAGTGTGATGATCTTCAAGCAGAGCCGCTGAT 393
DB 275 CTACATGATGAGACCGCCGCTGAGCAGTGTGATGATCTTCAAGCAGAGCCGCTGAT 334
QY 394 GCTGACATGCTGTGCGGAGTCTGATACGGGCGCTTCTTGTGACATGAGCGTCTT 453
DB 335 GCTGACATGCTGTGCGGAGTCTGATACGGGCGCTTCTTGTGACATGAGCGTCTT 394
QY 454 TGATGCTTACATGGAACCTGTGCTCCCGGAGACAGTCCAGCTCTTTCAGTGGAGAAC 513
DB 395 TGATGCTTACATGGAACCTGTGCTCCCGGAGACAGTCCAGCTCTTTCAGTGGAGAAC 454
QY 514 CCGGAGCCCTGTGTTGCACTGCTGTGATCATCTCCCAAGATGAATCTATCCCG 573
DB 455 CCGGAGCCCTGTGTTGCACTGCTGTGATCATCTCCCAAGATGAATCTATCCCG 514
QY 574 GGTCTACTGAGGCTCTGTGAGTCAAGACGCTTGTGAGTGTGGAGAGGCGCTGCG 633
DB 515 GGTCTACTGAGGCTCTGTGAGTCAAGACGCTTGTGAGTGTGGAGAGGCGCTGCG 574
QY 634 CTCTTACAGCCAGTGTGCTCAAGAGTGTGCTTTCACCTGACGTCCCTTAACC 693
DB 575 CTCTTACAGCCAGTGTGCTCAAGAGTGTGCTTTCACCTGACGTCCCTTAACC 634
QY 694 GGTCTGGAAGAGCCCTTCCCTCAACTGATCTGATGAGCTGTGGAGAGGCGCGGG 753
DB 635 GGTCTGGAAGAGCCCTTCCCTCAACTGATCTGATGAGCTGTGGAGAGGCGCGGG 694
QY 754 CGTGTTCCTGTCCGAGAAAGCAAAAGGAGATCTCATGATTGACAGTTCGATTTGAT 813
DB 695 CGTGTTCCTGTCCGAGAAAGCAAAAGGAGATCTCATGATTGACAGTTCGATTTGAT 754

QY 814 GGGGACAGCATGAGCAAAAACTCAAGAGAGAGCAACCTACTATGTATGAGGCTCAT 873
DB 755 GGGGACAGCATGAGCAAAAACTCAAGAGAGAGCAACCTACTATGTATGAGGCTCAT 814
QY 874 CTGCCAAAGCCAGCTGGAGATCTCAAGACATCTCAGTCTTCCCAAGGCGCTGACGA 933
DB 815 CTGCCAAAGCCAGCTGGAGATCTCAAGACATCTCAGTCTTCCCAAGGCGCTGACGA 874
QY 934 AGCTTACCTGCTGTGCGGATGAGAGACCTGGTCAAGGCGCTGAGGCGCGGAGCTTCCG 993
DB 875 AGCTTACCTGCTGTGCGGATGAGAGACCTGGTCAAGGCGCTGAGGCGCGGAGCTTCCG 934
QY 994 AATGATGATTTGCTGGATTTGGAATTTCTTCCCATCTTCAAGACCTGGTGCATACAT 1053
DB 935 AATGATGATTTGCTGGATTTGGAATTTCTTCCCATCTTCAAGACCTGGTGCATACAT 994
QY 1054 CACCCAGCCAGAGGATGGGTATGACACAGCTTTCACACAAATGCCAGGATGCCCTTAA 1113
DB 995 CACCCAGCCAGAGGATGGGTATGACACAGCTTTCACACAAATGCCAGGATGCCCTTAA 1054
QY 1114 TGTCTCCAGAGCTTGGGCGTGTCTTGTGCGCTTGAAGAGTTCTGACTTCAAGAAC 1173
DB 1055 TGTCTCCAGAGCTTGGGCGTGTCTTGTGCGCTTGAAGAGTTCTGACTTCAAGAAC 1114
QY 1174 CTGTGCGGATGCCAGTAATTTGCTGGGCTACCGCCACGTGATCTGAACAGAGAGAG 1233
DB 1115 CTGTGCGGATGCCAGTAATTTGCTGGGCTACCGCCACGTGATCTGAACAGAGAGAG 1174
QY 1234 AATCGTGTGCTGATCTTCTGTACCTGAGATGTCGCCAGCAATCCATCAAGAGAG 1293
DB 1175 AATCGTGTGCTGATCTTCTGTACCTGAGATGTCGCCAGCAATCCATCAAGAGAG 1234
QY 1294 TTGAGAGGCTTGTGCTGACACCTGTGTGACGCTTCTGATGATGCTTCTGA 1353
DB 1235 TTGAGAGGCTTGTGCTGACACCTGTGTGACGCTTCTGATGATGCTTCTGA 1294
QY 1354 GCTTGTCTTACATCTGTGAGCTTAACTACATCTGTGGGATCTACACTGAGTGTGAGT 1413
DB 1295 GCTTGTCTTACATCTGTGAGCTTAACTACATCTGTGGGATCTACACTGAGTGTGAGT 1354
QY 1414 TGTGTCCACAGCTGTCTCAAGCAGAGAGACTTGTGTGCTGATGATGATGATGAT 1473
DB 1355 TGTGTCCACAGCTGTCTCAAGCAGAGAGACTTGTGTGCTGATGATGATGATGAT 1414
QY 1474 ACTGGGAGACCTTATGTGAGCAGACATCTCCACAGAGTGAAGAGGATTTGCTTCTTC 1533
DB 1415 ACTGGGAGACCTTATGTGAGCAGACATCTCCACAGAGTGAAGAGGATTTGCTTCTTC 1474
QY 1534 TTTTCTTGTATCTTCTGTCTGTGGGAGACATCTTGTGCTGAGGCTGATTTAA 1593
DB 1475 TTTTCTTGTATCTTCTGTCTGTGGGAGACATCTTGTGCTGAGGCTGATTTAA 1534
QY 1594 GCAGCAGACAGTATCAGTGAATTTGATTCATAAACCCTCTGTCACATCTTCCCAATG 1653
DB 1535 GCAGCAGACAGTATCAGTGAATTTGATTCATAAACCCTCTGTCACATCTTCCCAATG 1594
QY 1654 GGGAGATGATCTTTCACAAAGAGCTCACAGCATTTTCCACAGAGATGGAATTTGAG 1713
DB 1595 GGGAGATGATCTTTCACAAAGAGCTCACAGCATTTTCCACAGAGATGGAATTTGAG 1654
QY 1714 CCTTGTGAGTTTCCCAATGAGATTCAGAGAGAGTGGGAAAGCAAGTGTGATGCTTACT 1773
DB 1655 CCTTGTGAGTTTCCCAAT - GGATTCAGAGAGAGTGGGAAAGCAAGTGTGATGCTTACT 1713
QY 1774 ATGAGCTTGACAT - ACAGCTATCGGTTATCAGAAATATGAACAAATCTCTG - ACATA 1831
DB 1714 ATGAGCTTGACATCAGAGCTATCGGTTATCAGAAATATGAACAAATCTCTGACATA 1773
QY 1832 AGAGCAAGCTCTTAACTCAAGAGTGTGCTGCTGATTTGAATTTCA - TTTCCCTTGG 1890
DB 1774 AGAGCAAGCTCTTAACTCAAGAGTGTGCTGCTGATTTGAATTTCACTTCCCTTGG 1833
QY 1891 CATTTTCCATTTACATAGAAACTTTGACCTGTGAACCTTGGCATCTGTTAATACTTAAA 1950

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: August 14, 2003, 10:39:13 ; Search time 507.533 Seconds
(without alignments)
10866.196 Million cell updates/sec

Title: US-09-645-078-1
Perfect score: 2043
Sequence: 1 gaattcattgttgggta.....tgggatacctataaaaaaa 2043

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 2552756 seqs, 1349719017 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_19Jun03:*

1: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1980.DAT:*
2: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1981.DAT:*
3: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1982.DAT:*
4: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1983.DAT:*
5: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1984.DAT:*
6: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1985.DAT:*
7: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1986.DAT:*
8: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1987.DAT:*
9: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1988.DAT:*
10: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1989.DAT:*
11: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1990.DAT:*
12: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1991.DAT:*
13: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1992.DAT:*
14: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1993.DAT:*
15: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1994.DAT:*
16: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1995.DAT:*
17: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1996.DAT:*
18: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1997.DAT:*
19: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1998.DAT:*
20: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA1999.DAT:*
21: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2000.DAT:*
22: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001A.DAT:*
23: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2001B.DAT:*
24: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2002.DAT:*
25: /SIDSL/gcgdata/geneseq/geneseqn-emb1/NA2003.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1856.8	90.9	1979	22	AAK94229 Human full-length
2	1813	88.7	2065	21	AAZ94211 Human transferase
3	1784.8	87.4	2032	20	AAZ20792 Human glycosyl sul
4	1262	61.8	1333	24	AAK16947 Human l-selectin s
5	763.6	37.4	877	22	AAK91803 Human cDNA 5'-end
6	763.6	37.4	877	22	AAK91921 Human cDNA clone r
7	743.2	36.4	1926	20	AAZ20793 Mouse glycosyl sul
8	547.2	26.8	2988	21	AAK76156 Human ORFX ORF1711

9	480.2	23.5	517	24	ABK54724 Human colon cancer
10	478.8	23.4	505	24	ABV89280 Human colon cancer
11	412.8	20.2	1647	24	AAZ24670 Human drug metabol
12	412.8	20.2	1694	22	AAD02700 Human glycosyl sul
13	412.8	20.2	2544	24	ABN89506 Human corneal N-ac
14	412.8	20.2	160552	22	AAD02697 Human glycosyl sul
15	398.4	19.5	2044	22	AAD02699 Human glycosyl sul
16	398.4	19.5	2170	22	AAD02698 Human glycosyl sul
17	387.2	19.0	1937	24	AAK16948 Murine intestinal-
18	387.2	19.0	1989	22	AAD02696 Mouse glycosyl sul
19	376.2	18.4	548	22	AAK92588 Human cDNA 3'-end
20	336.2	16.5	48436	24	ABN89533 Human corneal N-ac
21	294.4	14.4	389	24	ABK54794 Human colon cancer
22	144.8	7.1	7099	25	ACC46083 Human d1thp enzyme
23	141.6	6.9	2156	19	AAV21200 Glycosaminoglycan
24	141	6.9	668	24	ABO21506 Oligonucleotide fo
25	141	6.9	668	24	ABO21507 Oligonucleotide fo
26	135.2	6.6	1458	19	AAV36418 Keratan sulphate 6
27	135.2	6.6	2190	24	AAAD41280 Human chondroitin
28	135.2	6.6	2415	25	ABX08787 Angiogenesis-assoc
29	128.4	6.3	2354	18	AAT45037 Chick chondroitin
30	123.8	6.1	3029	22	AAH17922 Human cDNA sequenc
31	119.4	5.8	2393	24	ABK61699 Lung small cell ca
32	119.4	5.8	2409	20	AAK87821 Human N-acetylgluc
33	109.4	5.4	524	24	ABO21510 Oligonucleotide fo
34	109.4	5.4	524	24	ABO21511 Oligonucleotide fo
35	108	5.3	2150	20	AAK87820 Mouse N-acetylgluc
36	106.4	5.2	524	24	ABO21512 Oligonucleotide fo
37	106.4	5.2	524	24	ABO21513 Oligonucleotide fo
38	104.4	5.1	668	24	ABO21508 Oligonucleotide fo
39	104.4	5.1	668	24	ABO21509 Oligonucleotide fo
40	91	4.5	576	24	ABO45408 Oligonucleotide fo
41	91	4.5	576	24	ABO45409 Oligonucleotide fo
42	85.2	4.2	2165	25	ABK21794 Human chondroitin
43	68	3.3	576	24	ABO45406 Oligonucleotide fo
44	68	3.3	576	24	ABO45407 Oligonucleotide fo
45	62.6	3.1	731	20	AAZ24563 Human lung tumor a

ALIGNMENTS

RESULT 1	
AAK94229	
ID	AAK94229 strand; cDNA; 1979 BP.
AC	AAK94229;
DT	06-NOV-2001 (first entry)
DE	Human full-length cDNA, SEQ ID NO: 2816.
KW	Human; full length cDNA; cDNA synthesis; oligo-capping; ss.
OS	Homo sapiens.
PN	EP1130094-A2.
PD	05-SEP-2001.
PF	07-JUL-2000; 2000EP-0114089.
PR	08-JUL-1999; 99JP-0194486.
PR	11-JAN-2000; 2000JP-0118774.
PR	02-MAY-2000; 2000JP-0183765.
PA	(HELI-) HELIX RES INST.
PI	Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI	Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
DR	WPI, 2001-524255/58.
DR	P-PSDB; AAK93309.